

Putting a price on Maastricht

The failed summit meeting of the European Communities at Lisbon last weekend is another sign that Europe's most urgent need is better government.

THE supreme Council of Ministers of the European Communities (EC) at Lisbon last weekend was a predictable (and predicted) disappointment. It decided nothing of substance. More accurately, it confirmed that it would meet next at Edinburgh, towards the end of the six months during which the British will nominate the chairmen of other meetings of the council, to do with matters such as the working of the free market (due to start on 1 January next) and what the EC calls international relations. Meanwhile, the European enterprise will mark time, while it waits to see if last December's Maastricht Treaty is ratified. Nothing could have happened at Lisbon because the Danish government had been bound by a referendum last month to say NO. (Although Ireland has said YES in a similar referendum, that outcome is not binding on the Irish government — which will make the most of that freedom in the months ahead.) The first sign of real movement will not come until September, when there will be a referendum in France.

This is a crisis for the EC similar to, but more serious than, the crises it has wished on itself at intervals of about five years. It is therefore important that the crisis does not imply that the European enterprise as such is necessarily at risk. The Maastricht Treaty is an amendment of the Treaty of Rome, which established the European Economic Community. If one of the major founder members (France, Germany or Italy) were to fail to ratify, Maastricht would be a dead duck, but all would remain bound by Rome and, in particular, by the European Single Act from which next January's free market stems. (If Britain said NO, which it may well, the others would no doubt recommend secession.)

A few weeks ago (see *Nature* 357, 347; 1992), this journal was complaining that Maastricht is in trouble because it increases the executive power of central governments while doing nothing to make the process of government more democratic. It is also a fair complaint that governments had done too little to explain to their electors what Maastricht is about. Now, after Lisbon, it seems that they have not bothered to explain the treaty even to each other. Nothing else can account for the row at Lisbon about the EC's future budget, which has not been agreed.

One of the objectives of the treaty is to prepare the ground for a single European currency, to which end there is to be a European Central Bank directing monetary policy and, in advance, a period during which member governments contain their financial deficits within strict limits (3

per cent of Gross Domestic Product). Given that European exchange rates (with the exception of those of Portugal and Greece) are already virtually fixed within the Exchange Rate Mechanism, this is a recipe for keeping countries now relatively poor in perpetual relative poverty. That is why the Maastricht deal was to have been accompanied by the transfer of extra funds from the rich to the relatively poor countries. Ireland, Greece, Portugal and Spain would not have signed the treaty otherwise.

So why were the governments of Britain and Germany, while insouciantly pretending at Lisbon that Denmark has done nothing to shake confidence in Maastricht, also saying that they would not meet the extra cost (about £12 billion a year)? The charitable answer is that they have unexpected money problems of their own. Another is that they know that their chance of ratifying Maastricht will be enhanced if its costs are hidden from their own electors. It would have been reasonable to complain that the European Commission's calculations of what the transfers should be had been done on the back of some envelope, and that a more equitable and permanent basis of compensation for monetary union should be sought, but that is not what they said.

That is why the stalemate at Lisbon is yet another proof that, if Europe seeks closer self-association, it needs a much better way of governing itself. The British Foreign Office is boasting that it has arranged no fewer than 32 councils of ministers in the next six months, but nobody expects them to differ from earlier councils at which national ministers with sectional interests in, say, agriculture or research secretly barter compromises with like-minded but adversary sectionalists. Britain, for domestic reasons, is saying that there was a lot of talk at Lisbon of the need for more zealous application of the doctrine of subsidiarity (meaning the devolution of decision-making as far as possible down the administrative hierarchy), but nobody has yet explained how that squares with the European Commission's legal responsibility and power to remove (by often irksome regulations) national idiosyncracies held to impede international trade or threaten public health. In any case, subsidiarity seems not to apply to national governments' dealings with their own regional and local governments (except in Germany, where the *Länder* have constitutional rights). And there is hardly any talk of more power for a better European Parliament. There are altogether too many carts in front of too many horses. □



Let the French sue

A lawsuit against the US over the AIDS patent would finally bring all of the hard facts into open court.

THE case of *France v. United States*, or *Luc Montagnier v. Robert Gallo*, has been reopened by the French in the light of new data showing that the virus used in the US blood test is, as long thought, identical to the virus first discovered in France.

The dispute over the AIDS virus was first resolved in 1987 when a legally binding settlement over the rights to the patent for the AIDS blood test was signed by the French and US governments. Despite an agreement to share the credit and the royalties, and a legal promise not to reopen the case, the dispute has not come close to final resolution. This is demonstrated most vividly by a recent decision by US government lawyers to forbid a meeting at the National Institutes of Health (NIH) at which Gallo was going to be questioned about aspects of his laboratory's AIDS research (see page 3).

Underlying the US government's timidity is the fear that whatever Gallo said might undermine sensitive negotiations in Washington right now. The French are arguing privately that the US has a "moral duty" to give the French more than the agreed 50 per cent share of the patent royalties, now that it is clear that a previously unrecognized virus contaminated cell cultures in Montagnier's laboratory in 1983 and subsequently contaminated cells in Gallo's laboratory.

The French, represented by a New York law firm, have said that if there is no revision of the settlement agreement, they will take the US government to court. This should not be made or taken as a threat. If the French believe they can make a fair case for an increased share of the royalties, they should not hesitate to do so. If the United States believes that Gallo rightly earned 50 per cent of the credit for developing the patented blood test, it should not be reluctant to defend itself. (At present, the royalties, which come to about \$50 million, are split between the US government, the Pasteur Institute, and a new AIDS research foundation created as part of the 1987 settlement.)

For the sake of research and patients infected with the lethal human immunodeficiency virus that causes AIDS, this dispute must come to an end. Numerous attempts by various investigatory bodies, comprised largely of scientists, have failed to resolve points of contention about who said (or did not say) what to whom and when. Rumour and innuendo are carrying the day.

Scientists hesitate to go to court, but it could be that in this case the issues can be resolved only on the basis of hard facts that stand up in a court of law. Neither the NIH nor the US Congress should have any position of authority in evaluating the French case. Evidence should be brought out into the open; each party should cross-examine the other in court where all interested parties can observe the proceedings. It is now, perhaps, the only way people can ever be satisfied that justice has been done.

It often sounds pious to say that the real cost of scientific disputes is disruption of scientific research, possibly at the cost

of human lives. But in this case, the point may be fair. In Paris, four French health officials are on trial, charged with allowing AIDS contaminated blood to be distributed for five months after the US test was on the market (see page 6). The French government has acknowledged that some officials refused to approve the US test until a French test became available. More than 1,500 people (many of them haemophiliacs) are reported to be infected as a result.

There is no way to quantify the progress that might have been made if Montagnier and Gallo had been able to spend the past years fighting the AIDS virus instead of each other, but it is time for research to take precedence. And efforts to rewrite the settlement should not be made in private negotiations. □

Billion bites the dust

Nature intends that a billion should henceforth mean what others intend.

THIS journal hereby abandons a tradition that has served it well, if increasingly awkwardly, for close on 125 years. In this issue and from now on, the English word 'billion' will be understood as meaning 10^9 rather than 10^{12} . It is earnestly hoped that readers will not mistake this upheaval in editorial practice for the witless indulgence of innovation for its own sake. Rather, it is a case where tradition has been overwhelmed by others' usage.

For many years, it has been difficult to dragoon correspondents into describing, say, the US federal deficit as "close on \$500,000 million"; their base inclination is to say "\$500 billion", this journal's belief that even the US federal deficit cannot amount to $\$5.10^{14}$ in this or even the next presidency notwithstanding. Similarly, there have been difficulties over the age of the Earth, henceforth 4.5 billion years (but not, yet, '4.5 b.y.') More to the point, confusion about the meaning of 'billion' has created circumstances in which the word can hardly be used at all. To say that the US Gross National Product is a "few billion dollars" (which is true if one billion is 10^{12}) would seem demeaning to US taxpayers, most of whom consider themselves to be collectively three orders of magnitude better off. So the use of a valuable word has been sterilized and circumlocutions such as "500,000 million" have proliferated.

That is why, after a long struggle, it will in future be permissible for correspondents and contributors to use 'billion' in the sense of 10^9 . It is however, of the utmost importance that writers and would-be writers, whose frequent inventiveness in the use of words is often matched only by their indifference to the ugliness of what they thus do, should not regard this momentous change as a sign that the flood-gates are about to open. In particular, the word 'trillion' (for either 10^{12} or 10^{18}) will not be used except in direct quotations. Despite the permissiveness of Fowler's *Modern English Usage*, infinitives will not be split. And while nouns such as 'walk' are also verbs, not every noun can be dealt with in that way, as in "the sample was aliquoted . . .". There must, after all, be some standards. □

Robert Gallo forced off NIH stage after HHS vetoes public briefing

Washington. The US National Institutes of Health (NIH) were last week told to cancel a public meeting featuring Robert C. Gallo after officials of the Department of Health and Human Services (HHS) decided that such a meeting would be political suicide.

Observers say that HHS was concerned that an airing of the controversy surrounding Gallo's conduct of research over the past decade would disrupt delicate negotiations with the French government to resolve the question of royalty rights stemming from a blood test for the AIDS virus developed by Gallo and Luc Montagnier of the Pasteur Institute in Paris. Congressional pressure to delay the release of the report of a completed investigation into alleged misconduct by Gallo that has found him innocent of fraud but guilty of uncollegial behaviour may also have influenced the government.

On 23 June, Howard Temin, who shares a Nobel prize for the discovery of reverse transcriptase, received the US National Medal of Science from President George Bush in a ceremony in the White House Rose Garden. Less than two hours later, as chairman of the National Cancer Advisory Board's AIDS subcommittee, he received another, less welcome message from the government in the form of a hand-delivered letter from the general counsel of HHS, warning him to cancel the meeting.

Temin, outraged by the order, decided to hold the meeting just long enough to announce his feelings in public. "The forced cancellation does not give me a good feeling about the upper reaches of government", he says.

The offending letter was written by HHS counsel Michael Astrue. "This letter is to reiterate my oral advice to you, which is that the meeting you have scheduled for tomorrow to review matters relating to allegations of misconduct against Dr Robert Gallo exceeds the statutory authority of your committee and therefore must be canceled", Astrue wrote. "You need to be aware that the unauthorized expenditure of federal funds may expose you and others to various types of liability."

Temin admits to feeling threatened by Astrue's letter. "If I hadn't felt threatened I wouldn't have cancelled the meeting", he says. One week later, Temin says he still has no clear idea why the meeting was illegal and what liability he was exposed to. Furthermore, he calls Astrue's timing (less than 24 hours before the meeting) "insulting and rude".

Astrue says he was just applying the law. The cancer advisory board, which has oversight over the National Cancer Institute and

which is appointed by the White House, "has no authority to conduct misconduct investigations", he says. Temin's response is that "this would not have been a misconduct investigation. We were very careful about that."

Temin says the meeting was intended to "draw lessons" from the controversy, to ask questions about Gallo's management of his laboratory and to consider not fraud but



Howard Temin is outraged by the government's "rude" behaviour.

what he calls "sharp practices" in science — issues not of data but of collegiality and rudeness. "A society like the United States, with its savings and loan scandals and dirty politics, may not be one to expect that scientists and science should have no sharp practices."

Days before declaring the advisory board meeting illegal, Astrue also objected to it on policy grounds, saying that to allow Gallo to defend himself at a government meeting would "look as if the government had made up its mind in an ongoing decision process, when it hasn't". According to NIH sources, Astrue said that the meeting would be "purple Kool-Aid for the department", an apparent reference to the suicides by Kool-Aid of cult followers of Jim Jones in Guyana several years ago.

But Bernadine Healy, the NIH director who fought bitterly with Astrue in a failed attempt to salvage the meeting, does not accept that explanation. She says that HHS has "fundamentally criminalized a meeting of the National Cancer Advisory Board". Healy recently reviewed and accepted a

controversial report (two years in the making) from the NIH's now defunct Office of Scientific Integrity (OSI) that found Gallo innocent of fraud or serious misconduct in the AIDS virus affair (see *Nature* 357, 3; 1992). Healy's decision to accept the OSI's findings was powerfully influenced by the judgement of a committee of the NIH's scientific directors (many of whom openly dislike Gallo), who independently reviewed the report and concluded that the evidence is in his favour. To their own surprise, she says, "they agreed that he deserves the credit for developing the blood test".

Healy believes that Gallo, who has been the subject of numerous derogatory news articles and television shows, has been wronged and that the NIH has a duty to make things right. That is what lay behind her wish to hold the meeting. Healy, like members of the cancer board, wanted Gallo to have a chance to speak on the record, with the blessing of the NIH.

But it was not to be. Speculation about the reasons include the role of two other players who took an off-stage part in this melodrama.

New York attorneys for the Pasteur Institute, which is conducting secret negotiations with HHS for a larger share of the AIDS blood test royalties it now shares with the United States, sternly objected to the Gallo meeting and urged HHS to call it off because they saw it as a "press conference". Rebuffed at first, they then scheduled (and subsequently cancelled) a press conference of their own to follow the NIH meeting.

Then there is the question of the OSI report, still awaiting the signature of Healy's boss, James Mason, the HHS assistant secretary for health. For now, Mason has been effectively silenced by US Representative John Dingell (Democrat, Michigan), whose oversight committee has authority to review allegations of fraud. Dingell's staff recently sent Mason a 45-page rebuttal of the OSI report. Both documents have been leaked to the press.

The gist of Dingell's document is that the OSI report is faulty because of questions it allegedly failed to ask and conclusions it did not draw. This leaves Mason with a dilemma. He can support Healy's judgement and risk being called to explain himself before Dingell; he can order what would amount to a new study; or he can stall.

The latest word is that there will be no word for a couple of months at least, during which the melodrama will no doubt continue in the press.

Barbara J. Culliton

Italian agency questioned about first astronaut

Munich. Only weeks after the Italian government accused its space agency (ASI) of misappropriating research funds (see *Nature* 357, 351; 1992), the agency is being asked to defend its choice of the country's first astronaut.

Two weeks ago, Antonio Parlato, member of parliament for Naples, asked for a formal parliamentary investigation into the selection of Franco Malerba to fly on a joint mission with the United States to be launched later this month. Parlato alleges that Malerba did not qualify for selection because he was not scientifically qualified and he did not live and work in Italy. The mission, known as TSS, involves a satellite tethered to the space shuttle Columbus that will test the feasibility of obtaining power from space for a future space station.

In 1977, five candidates were chosen by Italy to fly with the first European space lab mission in 1983. None of them was chosen, but Guerriero contacted them again in 1983 when the National Aeronautics and Space Administration (NASA) offered Italy a slot on the TSS mission. The applicants were told that they must move to Rome and work on short-term contracts related to the mission. Three of them did so: Cristiano Cosmovici, Andrea Lorenzoni and Franco Rossitto.

The TSS was to be launched in 1990 and ASI began its selection process in February 1989. But the next month ASI readvertised on the grounds that the original candidates might no longer be medically fit. Fifteen new candidates were recruited and Umberto Guidoni, a plasma physicist at the Institute for Space Physics in Frascati, was selected as Malerba's backup.

Of the original candidates, Lorenzoni was eliminated by the scientific committee, and Cosmovici and Rossitto were ordered to undergo medical testing by the Italian air force, which declared both unfit. ASI's recommendation of Malerba was accepted by NASA's panel of experts, and he began his training in July 1989.

Malerba was working in Nice, France, in the sales department of the US-based Digital Equipment Company at the time of his selection and training. He did not receive a contract with ASI until November 1990. As such his appointment did not meet the requirement that the astronaut should be an experienced plasma physicist nor, according to Parlato, NASA regulations on employment. But Stephen Ballard of NASA says that "a relationship with a sponsoring organization is established when someone is nominated and they show up. Malerba's case is not exceptional."

Alison Abbott

Group asks NIH to stop growth hormone trials

Washington. A scientific watchdog group has petitioned the US National Institutes of Health (NIH) to halt a ten-year trial of human growth hormone in healthy children who are small in stature. Claiming that the experiment violates both the letter and the spirit of federal regulations controlling experiments on children, Jeremy Rifkin, president of the Washington-based Foundation on Economic Trends, says he will sue NIH if the trials are not ended.

US regulations state that federally funded trials involving children must either present no more than a minimal risk to the children or offer them the prospect of direct benefit. Exceptions are made for research likely to yield useful information about the children's condition.

NIH is conducting two trials, one in short children with normal levels of the growth hormone and a second involving children with Turner's Syndrome, a genetic disease that causes adults to be well under five feet tall. The syndrome affects only females.

In the case of the Turner's Syndrome trial, Rifkin opposes the use of a placebo-controlled group on the grounds that the trial (which involves ten years of weekly injections, regular X-rays and nude photography) is traumatic for the children. He points out that half of the 80 girls in the trial will eventually discover that they were actually members of the control group and had been denied a therapy that is effective only when begun early in childhood. Such use of children as "guinea pigs", Rifkin complains, is "exploitative and abusive".

He is even more upset by experiments involving children who are short but not deficient in the growth hormone. He cites studies that suggest human growth hormone therapy can produce side effects ranging from facial disfigurement to leukaemia and thyroid problems. Because being short is not a medically recognized disease or affliction, Rifkin says, the trials violate US regulations.

Many researchers have already made the same arguments and Rifkin's petition — filed on 22 June by his foundation and the Physicians Committee for Responsible Medicine — touches on an issue of considerable debate within the scientific community.

Several university Institutional Review Boards (IRBs) approved the Turner's Syndrome protocol in 1988, and some of those institutions now belong to the clinical network of institutions participating in the NIH trials. But several other IRBs, including that at the University of Nebraska,

rejected the protocol. Ernest Prentice, the assistant dean for research at the University of Nebraska Medical Center and the vice chairman of the Nebraska IRB, says "we felt that [the trial] was not only unethical, it was also not in compliance with the federal regulations".

Concerned that the regulations may be open to interpretation, "we did a great deal of research into the thinking that went into the regulations and their intent", Prentice says. "In this particular case it was pretty clear cut — there was just no adequate justification given" for going ahead with the trial.

The Nebraska group published an article arguing against the experiment in the January/February 1989 issue of *IRB*. Another group, from the University of Chicago, has argued (see *Journal of the American Medical Association* 261, 7; 1989) that short stature does not fit the usual definition of a 'disease' and that human growth hormone therapy for children not deficient in the hormone is of questionable benefit.

NIH officials believe that the two trials are necessary and appropriate. Arthur Levine, scientific director of the National Institute of Child Health and Human Development, says that some studies have shown that simply visiting a growth clinic and receiving a placebo makes some children grow. If so, he says, that would satisfy the federal regulations that such trials must offer a potential benefit to those in the control groups.

The Nebraska IRB rejects that claim, and points out that diabetic children are injected regularly and are not statistically taller than their cohort. (Levine disputes its logic, saying that diabetes itself affects height and could mask the growth-stimulating effects of injections.)

More important, Levine argues, is the fact that some doctors are already giving human growth hormone to healthy but short children. "Since it's going to be used, we have an obligation to learn if it is effective for the long term", he says. "Scientifically, we feel we're on solid ground."

Levine believes that such a research project satisfies the exemption in the regulations for trials that "present an opportunity to understand, prevent, or alleviate a serious problem affecting the health or welfare of children".

Rifkin's group has vowed to continue its opposition to the trials if NIH does not back down. "We've been fighting this for two years and we've exhausted our [administrative] remedies. We're ready to sue", he says, perhaps within a month.

Christopher Anderson

NIH makes room for the unconventional after Congress mandates new programme

Washington. The US National Institutes of Health (NIH) are putting the best possible face on an unwelcome order from Congress to spend \$2 million a year investigating unconventional medical practices. Insisting that it will maintain the highest scientific standards, NIH has created an office to assess the claims of practitioners of such unorthodox approaches as homeopathy, chelation and ozone therapy.

The order came last autumn in language accompanying the 1992 NIH budget written by the appropriations subcommittee chaired by Senator Thomas Harkin (Democrat, Iowa). It led, two weeks ago, to the first meeting of the *ad hoc* Advisory Panel on Unconventional Medical Practices, a group of 20 practitioners and supporters representing a broad array of therapies that fall outside the bounds of standard medicine. The group spent two days hearing from advocates of such practices and discussing the evaluation process. A second meeting is planned for the autumn that will focus on issues of methodology.

Even with an annual budget of \$9 billion, NIH would not have done such a thing on its own, admits Stephen Groft, the director and as yet sole member of the new office. "But money will force you into doing things that you otherwise might not do", he says. A few NIH institutes — notably cancer and infectious diseases — have in the past few years devised a procedure to look at unorthodox treatments, but NIH as a whole has been loath to embrace the idea that it is ignoring potentially life-saving interventions.

Hellen Gelband, who directed a 1990 study by the congressional Office of Technology Assessment (OTA) on unconventional cancer therapies, says that officials from the National Cancer Institute (NCI) treated OTA quite rudely in the course of its inquiry. "They told us they were doing everything that needed to be done, and that we should get out of their way", she recalls. Even before the OTA study was released, however, NCI began a programme that invites practitioners to submit their 'best-case series' of results for evaluation.

At the same time, the medical establishment is unhappy that Congress has once again set research priorities for NIH. Representatives of such groups as the American Cancer Society and the American Heart Association admit that some effective treatments were once considered beyond the pale and that it is reasonable to look at the most promising unorthodox approaches to combating cancer, heart disease, AIDS and other potentially fatal diseases. But they dislike the idea that NIH is being told to allocate limited research dollars in ways that might not do the most good.

"Like the poor, unconventional medical practices have always been with us, and there are treatments about which we will never have the answer", says Roy Schwarz, senior vice president for medical education



This 'colour therapy' lamp, seized from a California chiropractor in 1966, was said to cure some 30 conditions ranging from acne to asthma.

and science at the American Medical Association. "Although you hate to dismiss any potentially useful treatment out of hand, there's a lot of good science that isn't getting funded. We would hope that [the new office] consumes no more than a moderate amount of resources."

A congressional aide said that Harkin would like the office to have "half a dozen gumshoes, trained investigators who can go out and examine claims" of unorthodox medical cures and effective treatments. To that end, Groft has requested permission to advertise for three positions, to be filled by experienced researchers holding an MD or PhD degree and with knowledge of one or more unconventional practices.

For the programme to work, NIH must deal effectively with a group that mistrusts government and that has been a target of the US Food and Drug Administration. "If somebody from a federal agency raided my office, I'd be reluctant to show up on the doorstep of another agency", Groft says. "But our job is to listen to what they have to say."

But it takes two to communicate, and there are those who believe that NIH is

wasting its time. "The proponents of these therapies simply will not accept the results if they are negative", says Irving Lerner, a clinical professor of medicine at the University of Minnesota and author of a recent study by the Committee on Questionable Methods of Cancer Management of the American Cancer Society. "Just having the NIH look at it lends some credence to these questionable therapies."

On the other hand, says Berkley Bedell, a former member of Congress who advised Harkin on the 1992 report language, "the point is to find out what works. And I don't want some researcher to stand in the way of a cure for some dread disease just because it doesn't fit his definition of science."

NIH officials insist that they will not lower their scientific standards in dealing with this unorthodox community. "The more unconventional the approach, the more skeptically it will be viewed" explains Michael Friedman of the division of cancer treatment at NCI. At the same time, the new initiative will require NIH to be more open in judging the merits of a particular approach.

"Just as the zealot relies on his conviction that a particular treatment works, without collecting hard data, the scientific establishment has tended to dismiss these fanatics out of hand, without trying to find ways to obtain the necessary data", says Jack Killen, deputy director of the AIDS division at the National Institute of Allergies and Infectious Diseases. "Sometimes the only difference between unconventional and conventional therapies is the person using them."

Jeffrey Mervis

Panel trims NSF budget

Washington. The research budget of the National Science Foundation would remain level next year under a measure passed last week by an appropriations subcommittee of the US House of Representatives. The subcommittee ignored NSF's request for an increase of 13 per cent — an additional \$333 million — and instead gave most of the extra money that it had been allotted to veterans' health care and housing, two other programmes under its jurisdiction. "We're disappointed but not surprised", says one NSF official about the vote. "We knew all along that things were tight, but we were hoping for a little bit more than a flat budget." The next steps, votes by the full committee and House, are not expected much before the end of the month.

Jeffrey Mervis

Japan moves gingerly towards guidelines on gene therapy

Tokyo. Japan took its first step towards gene therapy last week when a committee of the Ministry of Health and Welfare (MHW) recommended the establishment of guidelines and the creation of a central committee to review research and clinical trials involving humans. But, although Japanese medical researchers are eager to follow the lead of the United States in this field, the Japanese must overcome fundamental weaknesses in the system of reviewing recombinant DNA research before they can begin gene therapy.

In bureaucratic Japan, there are almost as many guidelines and committees for recombinant DNA research as there are government ministries and agencies. The Ministry of Education, Science and Culture (MESC) has a committee to review proposals from universities, while the Science and Technology Agency (STA) has a parallel committee to deal with all other national research laboratories. On top of this, the health ministry, the Ministry of International Trade and Industry and the Ministry of Agriculture, Forestry and Fisheries each have guidelines for industrial applications that fall within their respective territories.

None of these committees or guidelines has so far addressed gene transfer in humans because the issue has been considered "taboo", says Mitsuru Miyata, editor of the influential newsletter *Nikkei Biotechnology*. But pressure from the medical research community and the pharmaceutical industry, combined with last week's announcement by the health ministry, is expected to spur action by other government agencies.

The government bodies dealing with health, science and education all play an important role in clinical applications of the technique. MHW oversees drug companies, hospitals and most medical research insti-

tutes, MESC is responsible for university hospitals and STA approves recombinant DNA research in all national research laboratories outside the universities. Despite their interest, it is not obvious what should happen next.

MHW officials say they have "no clear idea" of the form the proposed committee or guidelines should take and that their proposal is intended to "stimulate discussion". However, the MHW committee report calls for a group comparable to the Recombinant DNA Advisory Committee in the United States. Miyata and others say that MHW would expect to be in charge of establishing guidelines for commercial medical applications of gene therapy.

Quite apart from these bureaucratic turf wars, supporters of gene therapy must also cope with a lack of adequate systems for review of clinical applications of gene therapy at medical research institutions. Hospitals and medical research institutes have internal review boards to screen the safety, ethics and efficacy of new medical techniques. But the composition and effectiveness of these committees varies greatly from institution to institution, says Masanori Fukushima of Aichi Cancer Centre, a vocal critic of clinical research in Japan.

Some review boards contain experts from science and the law as well as members of the public. But many consist solely of members of the institute, who may not understand the role they are supposed to play, according to Fukushima, and who seldom stay in touch with medical experiments they have approved. Before guidelines for gene therapy are designed, he says, Japan needs rules for the establishment of institutional review boards.

There is also the question of why Japan

should do gene therapy at all. Medical researchers want to use it to catch up with the frontiers of science, but Fukushima warns that the scientific basis of gene therapy in Japan is "weak and thin" and that Japanese physicians "lack a clear strategy" for applying the technique.

The movement towards gene therapy highlights the problems that Japan faces in handling the ethical, social and legal issues arising from genetic research. The government invests a tiny amount in these issues compared with what is spent in the United States, Europe, Canada and Australia. Japan has few lawyers, ethicists or social scientists familiar with the rapidly evolving field, says Miyata, and those who do take an interest are "extremely conservative".

The health ministry intends to produce more detailed guidelines by the end of the year. In the meantime, the reactions of other government ministries and agencies to the preliminary plan is eagerly awaited.

David Swinbanks

French AIDS scandal goes to court

Four former French health officials went on trial last week in Paris on charges that they knowingly authorized the distribution of blood infected with the human immunodeficiency virus (HIV) in 1985. The scandal threatens to expand to other French political figures and could even touch on the continuing dispute between the United States and France over the discovery of HIV.

The officials are charged with allowing contaminated and unscreened blood to be used for five months after screening tests were available. A government investigation concluded last year that the officials delayed the use of a US blood test until a test manufactured by Diagnostic Pasteur, the commercial arm of the Pasteur Institute, could be produced (see *Nature* **353**, 197; 1991).

Lawyers for Robert Gallo, the US National Institutes of Health researcher who is in dispute with the Pasteur Institute over the discovery and patent ownership of the AIDS virus, say that the scandal shows that Pasteur scientists were not then as aware of the role of the AIDS virus as they now claim. François Gros, a former Pasteur director, was chairman of the meeting at which the decision was made to delay the test, and Luc Montagnier, the Pasteur researcher and co-discoverer of the virus, was also a government adviser. According to a transcript of an important meeting, French officials seem to have been advised wrongly that HIV did not appear to cause AIDS in most cases.

One of the officials now on trial, Michael Garretta, the former director of the national transfusion centre, has said that he and the other defendants are being made scapegoats and that the court should expand its investigation.

Christopher Anderson

US companies stand on the sidelines

Tokyo. Several small US companies are waiting for the Japanese government to approve applications of gene therapy (see above). They are landing on Japan's shores for much the same reason that the fledgling US biotechnology industry sought Japanese support in the early 1980s: they need money. But the country's large population and the high per-capita expenditures on medical treatment also make it an attractive market for US companies.

Among the leaders is Viagene, Inc. of San Diego, California, which is racing ahead in the development of a gene therapy treatment for AIDS in the United States (see *Nature* **357**, 615; 1992). In 1991, Midori Juji, a company that dominates the huge blood-product market in Japan, obtained worldwide marketing rights for Viagene's products for treating AIDS in exchange for funding the entire line of research at a cost of \$40 million over four years. The Viagene-Midori Juji combination is expected to be one of the first private-sector initiatives to develop gene therapy in Japan. But Midori Juji's goal of beginning clinical trials by the end of 1993 may be optimistic, given the time needed in Japan to create a system to review such clinical trials.

D.S.

Ban on primate trade seen as threat to animals

London. Primatologists are concerned that a new campaign by British anti-vivisectionists to end the captive breeding in source countries and international trade of primates used in research could worsen rather than improve conditions for the animals. The researchers fear that a ban might drive the business underground and shift it to countries less able to make sure that the animals are treated properly.



A Mauritian trapper holds a macaque caught in the wild.

Last week, the British Union for the Abolition of Vivisection (BUAV) announced that it would be putting pressure on countries involved in the export of wild-caught primates for research purposes. It singled out Indonesia, the Philippines and the island of Mauritius, all of which were visited by BUAV activists in the course of a year-long investigation. It also claimed that many of the captive breeding programmes it had seen were just a front for extensive trapping in the country concerned.

Groups such as the BUAV have a good record of success on the issue, as Asian countries have generally been responsive to demands that the trade in primates be reduced or stopped. India was one of the largest suppliers of rhesus monkeys in the world until 1978, when it banned their export. Bangladesh introduced a ban in 1979, and Malaysia in 1984.

But researchers are concerned that such bans, by reducing the sources of supply, may simply transfer the trade to countries

less able to meet demand and less experienced in caring for such primate populations. They cite the case of Malaysia, where a ban meant abandoning a small but well-regarded captive breeding programme.

One British primatologist who has spent much time in South-East Asia disputed BUAV's claims that captive-breeding programmes are just a front for wild trapping. In the Philippines, for instance, wild trapping is carried out under strict governmental control, in terms both of how many primates should be taken and from which areas. The Philippine government is committed to reducing the quota for the trapping of feral monkeys by 25 per cent each year while its captive breeding programme is expanded to meet demand.

People who have been involved in setting up supply lines for research primates point to sophisticated captive-breeding programmes in South-East Asia, some of which are of higher quality than those in more developed Western countries. Primatologists and the governments of source countries agree that high-quality captive breeding is the best way to meet the demand for research animals.

Even so, many British researchers and research institutions remain unconvinced that source countries can operate successful captive-breeding programmes. They are unhappy with the current piecemeal nature of regulations, and with the scarcity of documentation accompanying primates from source countries. The Medical Research Council insists that primates should come from within Britain wherever possible, except when a particular species is needed for which no breeding colony exists. The University of Oxford has its own breeding colony, chiefly so that it has control over the condition of the animals.

The BUAV will also be lobbying the British government to impose a ban on the import of wild-caught and captive-bred primates. Current laws allow the use of wild-caught primates in experiments only if there is no other source.

Nevertheless, the relevant legislative committee is concerned about the issue, and it is possible that it may extend the criteria for assessing suffering versus scientific benefit so that the suffering over the whole of the animal's life may be taken into account.

Ian Mundell

Outlook improves for Flemish researchers

Brussels. The Flemish government is trying to make amends to researchers after four years of harsh treatment. This summer, it is expected to ask its research council to reverse a controversial decision to abolish permanent research posts, after deciding in April to restore BFr60 million (US\$2 million) in research funds. The moves are part of a general government reorganization and reinstitution of policy planning for research — something that has not been possible since 1988, when Belgium was federalized and responsibility for basic research was placed in the hands of its two communities, Flemish-speaking Flanders and French-speaking Wallonia.

The initial government for the Flemish community, which constitutes 58 per cent of the population, did not have a system for protecting basic research. There was no minister with responsibility for science policy, and the prime minister could only juggle the various demands from individual members of parliament. In 1990, the government cut BFr40 million from its BFr1,960 million research budget, and this level of funding was maintained, without any allowance for inflation, for the next two years. The new budget, combined with an already low level of spending, resulted in no new scientific equipment and no additional permanent researchers. Some 5,000 scientists reacted to the worsening conditions by staging a pub-

lic protest in February (see *Nature* 355, 579; 1992).

The new government elected last November is much more sympathetic to the needs of research. The prime minister, Luc van den Brande, now has constitutional responsibility for science policy in Flanders. He has returned BFr60 million to researchers and pledged that future budgets will rise by at least the rate of inflation.

Van den Brande wants to restore the permanent positions immediately as part of an overall restructuring package.

Dirk Callaerts, who advises the prime minister on scientific matters, says that no final decision has been made on the restructuring plan and how permanent research positions may fit into it. "But unless we give researchers a future in Belgium we may lose them altogether", he says.

At present there are 203 permanent posts attached to university laboratories. Never intended to be truly permanent, the positions are used to employ the brightest researchers until they are offered a post within the university system itself. But the system has stagnated; knowing that the research posts will be lost if the current incumbent leaves, universities are loath to hire anyone. By contrast, in French-speaking Belgium, eight or nine permanent posts have been added each year since federalization.

Alison Abbott

NCAR under fire, a victim of its own success

Washington. The National Center for Atmospheric Research (NCAR) in Boulder, Colorado, may have become too successful for its own good. As scientists at the \$76-million research centre win more outside grants, the US atmospheric research community has become concerned that NCAR is abandoning its original mission to support their work. Last month, after independent reviewers criticized NCAR's management, the National Science Foundation (NSF) warned its overseers that it must change to preserve its relationship with the science agency.

At a meeting on 19 June, the NSF's National Science Board approved an agency decision to extend for one year rather than the usual five the contract with the University Consortium for Atmospheric Research (UCAR) to manage NCAR. NSF said the shorter extension resulted from questions raised during its most recent peer review of the centre.

Theoretically, NSF could 'recompete' the NCAR contract and choose a single company or university rather than UCAR's consortium of 59 universities. But most observers believe that the NSF intends its decision to be a statement to UCAR that it has stepped on the toes of too many university researchers and a suggestion that NCAR should grow more slowly.

On 19 June, Eugene Beierly, director of the NSF division of atmospheric sciences, wrote to Richard Anthes, UCAR president, to say that NSF and its peer reviewers had five major concerns: UCAR's management of NCAR, UCAR's reviews of NCAR's science, NCAR's funding from non-NSF sources, its plans for growth and the attention NCAR is paying to shared facilities, such as supercomputers and airplanes.

NSF did not release accompanying documents that elaborate on these issues. Nevertheless, UCAR officials and independent scientists familiar with the situation say that the letter reflects a growing unhappiness among university researchers over the following issues:

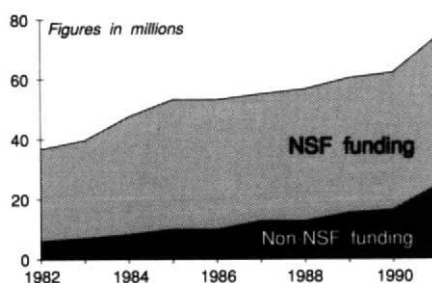
L Non-NSF funding: Since 1982, the proportion of NCAR's funding derived from sources other than the NSF has doubled, to a third of the total. UCAR officials say that Erich Bloch, the former NSF director, encouraged them to seek outside funding. But NCAR's success places it in direct competition with university researchers who vie for the same grants, and these scientists have made their displeasure known through direct complaints and barbed peer-review evaluation. "There's real resentment out here", says one academic scientist. "When NCAR competes for grant money, they're taking money out of the pockets of academic institutions."

Wayne Shiver, assistant to UCAR's

Anthes, says that "NCAR is doing what it was encouraged under one administration to do. Now, under a new administration, we're hearing something different."

L Facilities: NCAR was created more than 30 years ago mainly to provide a central resource for academic researchers who could not afford costly equipment such as

NCAR's growth



supercomputers and airplanes. Now that NCAR has its own thriving in-house research programme, outside researchers are concerned that it is not paying enough attention to shared facilities. With too little supercomputer time to go around, some atmospheric scientists want NCAR to put its money into new machines rather than its own research.

Shiver says that NCAR could never spend enough on facilities to satisfy all its users. Another supercomputer, for example, would cost an additional \$15 million a year, which

"is simply not in our budget". But NCAR officials are aware of the problem, and the centre has promised to spend whatever additional money it receives for 1993 on facilities rather than on research.

L Growth: Thanks in part to the rapid growth of both the NSF budget and the federal effort on climate change research, NCAR's budget has more than doubled during the past decade. Officials were concerned that such continued expansion would further strengthen NCAR's ability to compete for outside funding, which many non-NCAR researchers already see as a threat. They also feared that NCAR would forget about its core mission to support academic atmospheric researchers.

To address some of the criticism, UCAR intends to have teams from its member universities review specific programmes at NCAR and, possibly, ask NSF to help it select the reviewers and observe the process. "Initially the reaction was 'Oh God, the one-year extension sends a bad signal'", says Shiver. "But now there's a welcoming of the opportunity to resolve the issue."

Many researchers believe that NCAR needs more than fine-tuning. Some would like to see a number of regional centres, rather than one national facility. Others would like NSF to fund only the shared NCAR facilities, and let its researchers compete alongside their university colleagues. Says one researcher: "it may be time for NCAR, after 30 years of being nurtured, to stand on its own two feet".

Christopher Anderson

NEWS IN BRIEF

Washington. New rules that ease the tax burden on US high-technology companies with foreign operations were announced last week by the Internal Revenue Service (IRS). The change allows companies to deduct 64 per cent of their domestic research expenses, with the rest being taxed as if they were incurred abroad. US companies have lobbied hard for a change in the rule, adopted in the late 1970s to deal with the growing number of US companies with foreign operations. The rule turned out to be a disincentive for US-based research efforts, however, and a moratorium has blocked its implementation for most of the past 15 years. "We believe that the Bush administration has taken a giant step toward resolving a problem that has plagued America's research community for nearly two decades", says Roland Schmitt, chairman of the Council on Research and Technology, a coalition of universities, high-technology companies and trade associations that has worked to change the rule. **J.M.**

Washington. The US Congress failed last week to overturn a veto by President George Bush of a bill to end the ban on federally funded fetal tissue research. The 271:156 vote in the House of Representatives fell 14 votes short of the two-thirds majority required to override a presidential veto. But in an attempt to maintain pressure on the administration, Representative Henry Waxman (Democrat, California) on the same day introduced a bill that would require federally funded researchers hoping to do fetal tissue transplant research to apply to a bank that will collect tissue from ectopic pregnancies and spontaneous miscarriages. The bank was created last month after an executive order by the president. Under Waxman's bill, researchers would be free to seek the tissue from other sources if the tissue in the bank was not available or suitable. Waxman doubts that tissue from such sources other than abortion will meet the needs of researchers, but he believes that the only way to prove that point is to let the bank fail. **C.A.**

Internal politics block proposal by Russians to create foundation for basic research

Moscow. Two years of negotiations aimed at founding a Russian analogue of the US National Science Foundation (NSF) seem for the time being to have got nowhere. The plan, advocated by Boris Saltykov, the Russian minister of science, was to create a National Foundation for Fundamental Research (NFFR). But the scheme is now in danger of foundering over disputes about how it should be run.

The history of the foundation is instructive. It is two years since Igor Nikolaev, formerly a researcher, became an official of the Russian State Committee for Science and Higher Education with responsibility for the formation of Russian foundations. Persuaded that most pre-existing foundations were merely pots of money dispensed by administrative fiat, that supervisory expert committees were either nonexistent or were used to cloak administrative decisions in respectability and that the institution of peer review and the concept of conflict of interest do not exist in Russia, Nikolaev and two colleagues set out to draft the constitution of a Russian foundation that would function "properly".

After nearly a year and a half, they produced the papers for a foundation very similar to NSF. There was to be a scientific council, with authority divided between its president and a board of directors headed by a chief executive. Applications for funds would be scrutinized by committees of four experts.

To prevent the corruption of the appraisal process by the clannish Russian scientific community, expert appraisals were to be made public, referees were to be selected from a suitable database and foreign experts were to be involved. Nikolaev says the procedures for submitting applications and for accounting for funds spent were adaptations to Russian conditions of arrangements familiar in the West.

Towards the end of last year, the Carnegie Foundation of New York, which had seen the proposals, was talking of transferring a substantial sum of money to the foundation. But then the Soviet Union fell apart. When the wheels started turning again, and there was a presidential decree ready to sign, everything came to a halt again. The Ministry of Science (the post-*putsch* successor of the State Committee on Science and Higher Education) said there were so many pressing problems that the decree should be "more generalized".

So began another round of paper-chasing. The Carnegie money went elsewhere. The very idea of the foundation seemed to be an illusion . . . But then a miracle happened. On 27 April, President Boris Yeltsin signed a decree whose first item authorized the creation of the foundation as a self-governing entity. It also urged the recovery of Russian fundamental science and channelled 2 per cent of all Russia's spending on research to the foundation.

Sadly, the story does not end there. Yeltsin went on to appoint as president of the foundation Andrei Gonchar, the well-liked first vice-president of the Russian Academy of Sciences (RAS) and a natural candidate for the post. But when Gonchar held his first meeting with Nikolaev in mid-May, he made it plain that he sees the foundation in a very different light.

Gonchar said he did not fancy Nikolaev as executive director of the foundation and that there was no need for such a post. Nor was a scientific council necessary. Moreover, Gonchar objected to the condition that the presidency must be a full-time appointment, saying that he would not resign as vice-president of the RAS.

How this business will be decided is not yet clear. There is an obvious danger in the NFFR ending up as an adjunct of the RAS, which is responsible for only half of the fundamental research conducted in Russia. Or there may be another round of paper-chasing. The newly born foundation may become one of the best foundations in the world, but for now it exists only on paper.

Vladimir Pokrowsky

Italian peer review under pressure

Milan. Sandro Fontana this week becomes Italy's new minister for research at a time when most Italian scientists think the system funding them is unfair. One of the first demands on him will be a request from Luigi Rossi-Bernardi, president of the Italian research council (CNR), to change the rules governing the review of grant proposals to bring the CNR more into line with the rest of Europe. Responding to pressure from dissatisfied scientists, Rossi-Bernardi has agreed to ask the government to reconsider a system that 'doesn't know how to say no' — giving small grants to most applicants but large grants to no-one.

The CNR has ten committees that oversee funding in all areas of science. Each has around 15 members, most of whom are university professors elected every four years by the scientific community. Although democratic, the system does not produce a formal peer review system within the committees. Elected members feel a responsibility to their voters, and therefore award money on the basis of who holds the power within a university as well as on scientific merit. Rossi-Bernardi says that he would like to preserve the democratic element of committee elections while broadening the pool of reviewers.

Alison Abbott

Indian court bans quotas that give share of science posts to outcasts

New Delhi. An Indian high court has ruled that reserving positions for socially backward castes in government jobs and university science and medical programmes is illegal and a violation of the Indian constitution. The judgement by the Allahabad high court of Uttar Pradesh, India's biggest state, overturns existing policy that sets aside 22 per cent of the jobs and university admissions slots to those in the lower castes. The judge declared that science has no religion or caste and that "there can therefore be no valid reservation in the field of science and technology".

The current policy tries to balance the need for technical excellence with the obligation to make special provisions for those from educationally and socially disadvantaged classes. Although he said that backward classes must be helped, the judge ruled

that the interests of the nation are paramount. "The time has come when it must be boldly and clearly stated that there can be no compromise in the field of science and technology", he wrote in his decision.

The ruling is a response to a suit filed by a college chemistry lecturer from an upper caste who was ordered to vacate his position to make room for someone from a lower caste. The court invited the lecturer to remain and asked the college to pay his salary.

The caste issue is controversial in India. Two years ago, a suggestion by the prime minister that the quota be raised to 40 per cent triggered riots by students in which 35 people died. Although the ruling is not binding on other states, it is expected to shape judicial decisions throughout India.

K.S. Jayaraman

Questions about AIDS

SIR — A growing wave of popular press attention is currently being focused on P. H. Duesberg and the contention that the human immunodeficiency virus (HIV) does not cause AIDS. Two years ago, Duesberg listed 11 phenomena that he considered to be paradoxical in the context of the hypothesis that HIV causes AIDS¹. I responded by showing how these paradoxes can be resolved in the context of an autoimmunity model of immunopathogenesis². Duesberg has not responded to these resolutions of his paradoxes. The set of paradoxes is potentially useful. The ability to resolve them is a convenient benchmark for the evaluation of competing theories of AIDS pathogenesis, and may bring focus to an otherwise rather undisciplined debate. I challenge proponents of other pathogenesis models to show that their models can match the performance of our autoimmunity model in resolving the Duesberg paradoxes.

Geoffrey W. Hoffmann

Departments of Physics and Microbiology,
University of British Columbia,
Vancouver, BC, Canada, V6T 1Z1

1. Duesberg, P. H., *Res. Immunol.* **141**, 5–11 (1990).

2. Hoffmann, G. W. *Res. Immunol.* **141**, 701–709 (1990).

SIR — Perhaps others in my position have thought about this question and would also welcome an answer. I receive a reasonably large number of reprints from areas of the world where AIDS is rampant, and would like some assurance that if one gets a paper-cut finger from the saliva-licked sealed envelope, one needn't worry about the possibility of viable AIDS virus passage. Is there some assurance/information knowledgeable readers can provide?

Lee Frank

Pulmonary Research Division,
University of Miami School of Medicine,
PO Box 016960,
Miami, Florida 33101, USA

SIR — At the alternative AIDS conference in Amsterdam in May, I proposed that American/European AIDS diseases, above their normal background, are caused by recreational drugs and the antiviral drug AZT. In a News and Views article (*Nature* **357**, 188; 1992), you blame me for “failure . . . to respond to the counterarguments”, reliance on “circumstantial” evidence, and for “enjoying [myself] so hugely at the expense of lesser mortals who have managed so grossly to misunderstand AIDS . . .”.

Unfortunately *Nature* fails to understand how difficult it is to present in one talk sufficient evidence to counter all claims of the virus-AIDS hypothesis that

have been publicized in more than 60,000 papers and how extremely difficult it is to publish the basis of my drug-AIDS hypothesis in the *Proceedings of the National Academy of Sciences (USA)*. My paper was rejected even though I belong to the academy and membership typically includes the privilege of publishing without editorial review.

I decided to soften my “punches” against the virus-AIDS hypothesis with humour. This is not to say that I enjoyed myself at the expense of lesser mortals. Indeed, the proponents of the virus-AIDS hypothesis are by no means “lesser”. Instead, I am advancing my hypothesis very much at my own expense. Since I challenged the virus-AIDS hypothesis, which is entirely unproductive in terms of public health benefits, I have been excommunicated by the retrovirus-AIDS community with noninvitations to meetings, noncitations in the literature and nonrenewals of my research grants, which is the highest price an experimental scientist can pay for his convictions.

Peter Duesberg

Department of Molecular
and Cell Biology,
University of California,
Berkeley, California 94720, USA

Why cheat?

SIR — The present debate over fraud and unethical behaviour in science (*Nature* **356**, 730; 1992) fails to ask why people cheat. Until this question is asked and answered, and the incentives for cheating are removed, some people will continue to do so, regardless of the rules and guidelines.

People cheat only if their expectation of gain exceeds their fear of exposure. Cheating does not advance science, so its presence indicates that something is wrong with the reward system. People are rewarded (by jobs or grants) for filling their *curricula vitae* with unexamined papers. This reward system is also responsible for such phenomena as the ‘minimum publishable unit’ and competitive ‘grantsmanship’, by which the efforts of honest scientists are diverted from productive research.

The present funding system in science (at least in the United States) is destructive because it has replaced healthy competition (to make discoveries or to develop useful products) with wasteful competition (to have the longest *c.v.* or highest ranked funding proposal). Cheating is rarely caught, and is rewarded in this wasteful competition. The

solution must restructure the evaluation process so that it is based on a proposer's entire record of accomplishment rather than on the promises made in a proposal. Reform should also guarantee a minimal level of support to every productive and original scientist, drawing the required resources from larger groups organized around a successful grant-getter. Such a concentration of people and resources into large groups chokes original and venturesome research because only the group leader is independent, and even he is constrained by the need to market his funding proposals to his peers and the pressure to find support for his group.

Competitive review of proposed research ensure that only consensus science will be supported. New and original ideas are unlikely to attract consensus approval, and are discouraged by the present system. Most scientists therefore disguise their intent in the proposals, which encourages more serious cheating.

Jonathan Katz

Washington University,
Campus Box 1105,
One Brookings Drive,
St Louis, Missouri 63130-4899, USA

Human insulin

SIR — In his interesting Commentary (*Nature* **356**, 375; 1992), Simon Wolff says that there is no controversy over human insulin in the United States. This is not the case; we are aware of several diabetic patients in the United States who are unhappy with human insulin, and there are legal cases in process now where doctors are being sued for continuing to prescribe it. There are diabetic patients in other countries who have had problems with human insulin, but where the authorities do not admit there is a problem. These include France and notably Australia, where all other forms of insulin have been withdrawn. It is our view that the only reason there are no acknowledged problems with human insulin in these countries is that they do not have a method of collecting the information or that patients are not being listened to.

Wolff does not mention that a number of diabetic patients have been changed back from human insulin to porcine insulin, and that these patients find that their symptoms of hypoglycaemia have returned. No account has been taken of this fact by anybody, and until someone is able to collate this information it may be that we are still sitting on a human insulin timebomb.

M. R. Kiln

12 Elm Road, Beckenham, Kent, UK

Amanda Sugarman

12 Primley Park Walk, Leeds, UK

Trial and error in the Highlands

T. H. Clutton-Brock and S. D. Albon

How should the increasing numbers of deer throughout the world be controlled? The situation in Scotland shows that an understanding of the biology of deer can provide practical guidelines.

ALTHOUGH numbers of many large mammals have declined over the past century, deer populations throughout much of the Northern Hemisphere have shown a dramatic increase¹. Linnaeus is supposed never to have seen a wild moose, but in Sweden today there are more than 300,000 of these animals (*Alces alces*) and more than 100,000 are shot each year. The numbers of American white-tailed deer (*Odocoileus virginianus*) are thought to have risen from less than 500,000 at the turn of the century to 20 million by the mid-1980s¹. The causes of the increase vary between species and country, but include the abandonment of marginal agricultural land and associated increases in natural woodland or commercial forestry, reduced competition with domestic stock, and more effective controls on hunting.

Increasing deer numbers have permitted larger harvests to be taken each year. In 1985, close to 6.7 million animals were killed from populations totalling at least 37 million¹ in Europe and North America. Though the economic value of hunting is unknown, it is clearly substantial; in most countries in Europe large areas of upland are managed partly or exclusively for hunting. In 1976, around \$2,600 million was spent by white-tail and mule deer hunters in the United States, equivalent to around \$1,000 per animal killed.

Ecological damage

Various ecological problems have arisen as deer populations have spread and their numbers have approached carrying capacity, including degradation of habitat, damage to forestry or native

woodland, an increase in the incidence of arthropod-borne diseases (especially Lyme's disease which is more common in the United States and is spreading rapidly in Europe) and, in Sweden, an increase in deer-related traffic accidents.

Where predators are rare or absent, it can be necessary to control deer numbers in advance of the natural limits to density imposed by food availability, parasites and disease. Where game is treated as public property, as in the United States and Canada, state wildlife agencies have had some success in controlling deer populations by varying the length of close seasons and the number of licences given for different age and sex categories of deer¹. But where the right to kill game is vested in private landowners, as in many European countries, conflicts of interest are likely to arise between owners wishing to maximize annual offtake from their ground and conservation agencies responsible for preserving the habitat.

In Scotland, conservation agencies are now calling for statutory controls on red deer (*Cervus elaphus*) density. The case is interesting in that it demonstrates how an understanding of the biology of deer populations is necessary before legislation can be sensibly introduced. Red deer are indigenous to Britain but by 1800 their numbers had been reduced to low levels. In the first half of the nineteenth century the growing demand for sport hunting (stalking) by English gentry persuaded Scottish Highland lairds to conserve their remaining deer populations: the area of ground managed primarily for deer in Scotland (named deer 'forests' despite the lack of

trees) rose rapidly to nearly 2 million ha by the turn of the century when more than 10,000 deer were being killed annually. The first systematic counts between 1952 and 1960 suggested a total population of around 155,000. Since then, numbers rose to 200,000 by 1974, 255,000 by 1979, 290,000 by 1986 (ref. 2) and now stand at around 300,000.

Size increase

Recent increases in population size have been caused partly by an extension of deer populations into new areas and partly by an increase in deer density over much of their established range (Fig. 1a). Although it is often suggested that increased density *in situ* has been caused by a decline in culling, the total number of animals killed shows no consistent pattern³. Most of the 16,000 stags killed in the Highlands each year are shot by clients or tenants who pay around £200 to shoot each animal, and both the income and the capital value of deer forests are closely related to the number of stags that can be shot. By contrast, hinds are culled, often at a net loss, by estate employees between mid-October and mid-February when weather conditions make it difficult to hunt animals.

It is now widely argued that there are too many deer in the Highlands^{2,6}.

The Red Deer Commission, the statutory body responsible for monitoring Scottish red deer populations, is now calling for a 30% reduction in hind numbers²; the World Wide Fund for Nature is being quoted as being in favour of a 50% reduction; and Scottish Natural Heritage (the government's agency) would like to see numbers reduced to a level that would allow unfenced natural woodlands to regenerate⁷. The British government's Scottish Office is now seeking views on whether it should empower Scottish Natural Heritage to compel landowners to reduce deer numbers if there is "serious damage to the natural heritage interest"⁸.

There is certainly a case for reducing Scottish deer numbers but many of the arguments for doing so have been exaggerated. In addition, recent increases in deer numbers have probably been caused partly by an increase in the number of deer the ground can support, arising from a reduction in competition with sheep². Numbers of ewes in the Highlands of Scotland fell by around

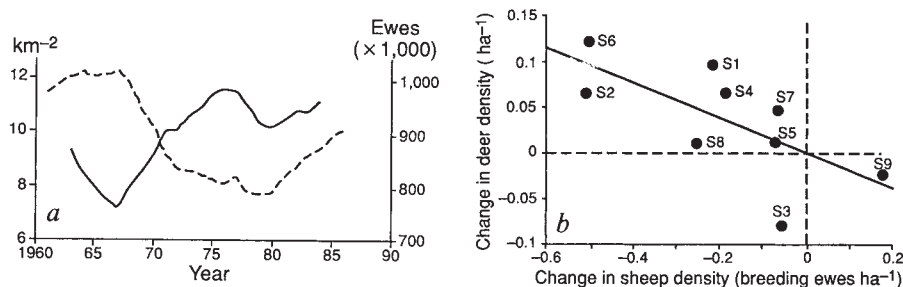


FIG. 1a, Changes in average red deer density (solid line) throughout the Highlands and in the number of breeding ewes using the deer range (dashed line) since 1960 (from ref. 3). Estimates of deer density exclude deer in forestry plantations and are based on ground counted at least twice by the Red Deer Commission over this period. b, Changes in deer density on estates which made substantial changes to their sheep stocks and for which deer counts were available before and after the change in sheep numbers. $y = -0.2x$, $x = -0.0008$, $r^2 = 0.43$, $P < 0.05$.

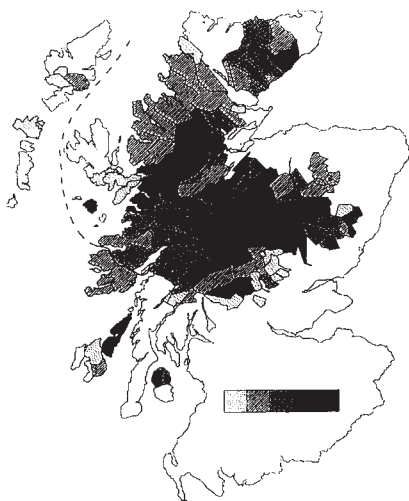


FIG. 2. Total deer density map for parishes covered by the Red Deer Commission's counts, standardized for 1986 (from ref. 3). Broken line, boundary of the highland region. Some unmarked areas (for example, northwest Sutherland) contain red deer stocks that have yet to be counted. Darkest areas, > 20 deer per km²; lightest areas, < 5 deer per km².

20% between 1966 and 1978 when deer numbers rose (Fig. 1a) and removal of sheep on particular estates is followed by increases in deer numbers (Fig. 1b), indicating that the two species compete for resources.

The best arguments for reducing Highland deer numbers are based on the need to stop long-term changes in upland plant communities. Several upland plant communities are susceptible to grazing by deer and at least three of these have incurred more than 2% loss since 1954 (ref. 7). Many natural woodlands used by deer show little regeneration because of browsing^{3,6,7}. Since 1945 there has also been a gradual loss of dwarf shrubs in some areas, probably caused in part by the impact of herbivores⁷. Unfortunately, it is seldom possible to be certain to what extent deer versus other factors (including the impact of sheep and the burning of moorland vegetation in early spring) have been responsible for these changes or to determine the effects of recent changes in deer numbers. In addition, it is not clear how much deer (and sheep) densities would have to be reduced. Finally, a drastic reduction in deer numbers could have far reaching consequences. If the interests of estate owners were disregarded and deer numbers were reduced to low levels throughout the Highlands, this would almost certainly lead to increased pressure for land to be used for other purposes, such as commercial forestry and skiing, likely to have more detrimental effects on natural plant and animal communities.

What is to be done? There is an urgent need for large-scale experiments to investigate the effects of different

densities of sheep and deer on upland vegetation and to determine the approximate density of herbivores that different plant communities can sustain. It is less clear that the government should attempt to limit deer numbers by legislation, for the carrying capacity of hill ground varies widely (deer densities vary from less than 5 to more than 25 per km² between parts of the highlands; Fig. 2) and cannot be reliably predicted. Nor will it be straightforward to define what constitutes "serious damage to the natural heritage" or, where damage has occurred, to determine to what extent deer are responsible. In many areas of the western Highlands, there would be little point in limiting deer density if the density of sheep (six times as numerous as deer) is not controlled too.

Statutory limits on deer and sheep densities may be necessary in some areas but should be a last resort. What other measures might help? It would be sensible to advance the start of the hind shooting season to the middle of September, when animals are more accessible and retrieval is easier than in midwinter, as is done in several other European countries. But a more pressing need is to convince landowners that it is in their own interests to reduce their hind populations. This should not be impossible. Where populations of deer (and most other ungulates) are managed primarily for offtake of mature males, it is usually necessary to maintain female numbers well below the maximum level that the environment can sustain because high female densities and low food availability depress the growth and survival of males more than females^{3,10}. This is probably partly because sexual selection for large adult size in males has raised their growth rates and energy requirements above the ecological optimum, and partly because males deplete their resources during the annual rut so are more susceptible to lack of food in winter than females. As a result, many ungulate populations that are subject to little or no culling or predation show a strongly female-biased adult sex ratio³ and poor growth in males.

Differences in the response of males and females to low food availability have important implications for management strategies. Where females are lightly culled and predominate in the adult population, managers should be able to increase both the number and the quality of mature males taken each year by raising the rate of female culling. In highland red deer populations where less than 6% of breeding females are culled each year, for example, managers may be able to increase their offtake of 5-year-old stags by 30% or more by raising hind culls to 16–18% (Fig. 3). Because this increase in yield could have

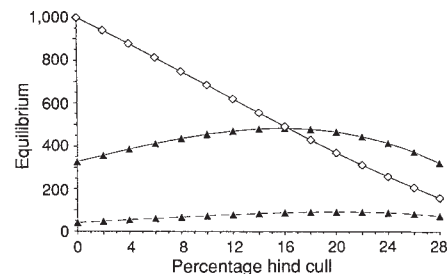


FIG. 3. Likely effects of varying the rate at which females (diamonds) are culled on numbers of males (triangles) ≤ 1 year old (solid line) and ≥ 5 years old (dashed line) in a simulated population of 1,000 hinds. The maximum cull of mature (≥ 5 -year-old) males is achieved by removing all animals of this age or above each year. Estimates of sex-specific changes in survival, fecundity and emigration based on density-related changes in the red deer population of Rum³.

a substantial effect on both the income and the capital value of estates, it ought to be a telling argument.

A need to control deer numbers (and female numbers in particular) is likely to develop in many areas of Europe over the next decade. At the root of the problem lies the preoccupation of hunters with trophies and, consequently, with shooting males. This practice is less prevalent in North America, partly because hunting pressure is commonly high and hunters are less concerned about the age and sex categories of their prey whereas, in Europe, relatively little income is derived from hunting females and a substantial industry revolves around male trophies. Perhaps the hunting lobby and conservation agencies should combine to stimulate hunters' interest in females. Who will offer a prize for the smallest red deer hind shot in the Highlands this year? □

T. H. Clutton-Brock is in the Department of Zoology, University of Cambridge, Cambridge CB2 3EJ; S. D. Albon is at the Institute of Zoology, London NW1 4RY, UK.

1. Gill, R.M.A. *Monitoring the Status of European and North American Cervids* (Final Report for UNEP/GEMS Project ST/4101-84-02, UNEP 1988).
2. The Red Deer Commission *Annual Report for 1989* (HMSO, Edinburgh, 1990).
3. Clutton-Brock, T.H. & Albon, S.D. *Red Deer in the Highlands* (Blackwell Scientific, Oxford, 1989).
4. Albon, S.D. & Clutton-Brock, T.H. in *Ecological Changes in the Highlands* (eds Usher, M.B. & Thompson, D.B.A.) 93–107 (Blackwell Scientific, Oxford, 1988).
5. Callander, R. *The Management of Wild Red Deer in Scotland* (Rural Forum, Perth, 1991).
6. Mitchell, B., Staines, B.W. & Welch, D. *Ecology of Red Deer* (Inst. Terrestrial Ecol., Cambridge, 1977).
7. Francis, J.M., Balharry, R. & Thompson, D.B.A. *Deer, Mountains and Man* (NCC, Scotland, 1990).
8. Scottish Office Agricultural and Fisheries Department. *Consultation paper on Deer Legislation in Scotland* (1993).
9. Clutton-Brock, T.H., Albon, S.D. & Guinness, F.E. *Nature* **313**, 131–133 (1985).
10. Clutton-Brock, T.H. & Loneragan, M. *J. appl. Ecol.* (submitted).

ACKNOWLEDGEMENTS. We thank Scottish Natural Heritage, the Red Deer Commission, the British Deer Society for support, and M. Loneragan for permission to reproduce Fig. 3.

Media make AIDS wishes come true

A fuss last week about a crop of HIV infections in Birmingham (England) shows British newspapers, but notably the *Sunday Times*, at their most fickle.

FALSE prophets are well-known to be immune to reason and are even able to shrug off evidence that their prophecies are false, usually by playing on the wishes of their followers that the truth were otherwise. So it seems to be with the motley crew of British newspapers which, having first heard of Dr Peter Duesberg at the "alternative" conference on AIDS at Amsterdam a month ago (see *Nature* 357, 188; 1992), have since embraced his view that the human immunodeficiency virus (HIV) is an epiphenomenon irrelevant to the causation of AIDS and the comforting corollary that people have nothing to fear from heterosexual intercourse, provided that they do not take drugs or otherwise abuse themselves with chemicals.

Various tabloid newspapers have taken up this dangerous litany, but the most surprising convert to Duesberg's cause is the *Sunday Times*. (For a time, it seemed that its editorially independent stablemate *The Times* was about to follow suit, but that seems mercifully to have been an aberration; earlier this week, the newspaper was writing about the prospects for a vaccine against HIV based on the use of the pigtail macaque as an animal model of the infection.) Over the past few weeks, the *Sunday Times* has reported at length on the spurious controversy stirred up by Duesberg, has given him and his supporters a platform from which to declaim and has dealt dismissively with objections from people such as Sir George Calman, Chief Medical Officer at the UK Department of Health.

Why a serious newspaper should be following such a line is anybody's guess. In its time, specifically that of a previous editor, Mr Harold Evans, it won an excellent reputation for investigative reporting. It has also successfully championed important public causes, notably that the drug thalidomide was negligently put on the British market. But it is also known for having begun serializing 'Hitler's diaries' until they were shown to be forged and, more recently, to have bought the serial rights to a book claiming that the marriage of the Prince and Princess of Wales is in trouble, whose publication can only make matters worse.

But the *Sunday Times*'s prophecy on HIV and the likelihood that AIDS will spread heterosexually seemed last week to have been denied by a counter-example from Birmingham, where it emerged that a single male carrier of HIV — a haemophiliac infected by contaminated factor VIII — had in turn infected at least four women partners, one of whom had died of AIDS. (The information

appears originally to have been leaked to the *Birmingham Mail* by a physician frustrated by the lack of legal restraints on the carrier's activities.)

It would have been interesting to see how the *Sunday Times* would have squared these reports with its earlier stance on HIV, but in the event, it escaped the dilemma. By last weekend, the tabloid newspapers had worked out that the Birmingham haemophiliac had engaged in anal intercourse with at least three of the four whom he had infected, and the *Sunday Times* was able to proclaim on its front page that "NEW EVIDENCE CASTS DOUBT ON BIRMINGHAM AIDS SCARE". The implication was that unnatural intercourse was responsible for the Birmingham episode. Really straight heterosexuals might relax.

This really is a shameful way to carry on. Even granted that newspapers seek self-consistency, there is no justification for putting consistency before correctness in this way. And the plain truth is that there is no justification for the suggestion that the extra hazards of anal intercourse are novel. The most recent evidence to this effect is that published three months ago by the European Study Group on Heterosexual Transmission of HIV, which estimated that the relative risks of HIV transmission (from men to women) by anal and vaginal intercourse are in the ratio 5.1:1.0 (*Brit. med. J.* 304, 811; 1992).

Earlier estimates of the relative risk are lower. For what it is worth, the European study also showed that the overall risk of transmission from men to women (at 20 per cent) is twice that from women to men. Risk factors (for infection by HIV) other than anal intercourse include intercourse during menstruation and whether one or other partner is in a late stage of AIDS. Because the figures are based on a survey of partnerships lasting, in some cases, for several years, they cannot be directly used to derive the risk that a single act of sexual intercourse will be infective, which epidemiologists will have to continue to estimate by consistency with studies such as this.

None of this is at odds with the proposition (which Duesberg and the *Sunday Times* have taken to calling the "AIDS hypothesis"), that HIV infection is the precipitating cause of AIDS. On that view, the infective routes are sexual intercourse and direct contamination of the blood by transfusion or injection. Evidently, the chance of infection in a single act of intercourse is small, probably less than the chance of catching influ-

enza from an infected partner on a similar occasion. It is no scandal that estimates of the risks are still so uncertain: they can be derived accurately only from a detailed knowledge of the sexual behaviour of infected people, likely to be as reticent as others on that score.

Similarly, there are many other features of the development of AIDS that remain poorly understood. Thus it seems now to be widely accepted that the pathogenesis of AIDS is not yet understood. Infection by HIV does not quickly lead to the infection of each susceptible cell, let alone to their rapid destruction. Instead, immune deficiency may take years to develop. That is why people are looking for mechanisms to account for this prolonged response, auto-immune reactions perhaps. (The mere suggestion last year — see *Nature* 353, 297; 1991 — that such an explanation might get Duesberg off the hook on which he had impaled himself has been wrongly interpreted as support for everything he had been saying.) And who will deny that an understanding of the pathogenesis of AIDS after HIV infection is likely to provide the surest route to therapy?

Not, it seems, the *Sunday Times*. Why is it so perverse? The charitable explanation is that the newspaper is responding to the deep feelings of those infected with HIV, who understandably rage at the 'death sentence' they have supposedly been given. But there is a wider constituency, that of those who wish that heterosexual intercourse, however unguarded, has been given the relatively clean bill of health it enjoyed in, say, the 1960s. Even the newspaper that shouldered the thalidomide cause should give some thought to the huge responsibility that it is, as on this occasion, wrong.

There have already been other unwelcome responses to the Birmingham affair. The most sinister development is the demand that AIDS should be made a notifiable disease, which would bring into play the legal powers that allow the compulsory tracing of sexual partners (for venereal diseases) and even quarantine — an administrative recipe which would do more harm (by discouraging voluntary testing) than good. And there is the inevitable resurgence of homophobia, made the more vivid by the smug wish-fulfilment that no harm will come to heterosexuals from AIDS because they do only what they have always done, and also behave 'naturally'. Sadly, the novelty is HIV and the prophecy that heterosexuals have nothing to fear from it is false.

John Maddox

Beating the code breakers

Artur Ekert

QUANTUM cryptography — the application of quantum mechanics to assure secrecy of communications — is now taking off and can no longer be regarded as science fiction. In one new development¹, Charles Bennett of IBM Research shows how interference of very dim light pulses can be used in practice for what cryptologists call key distribution, any technique that allows two parties to acquire the same random sequence of numbers, the key, with a high level of confidence that no one else knows it. The key itself, a string of ones and noughts, is random and meaningless, but it allows enciphered communication to follow with provable security.

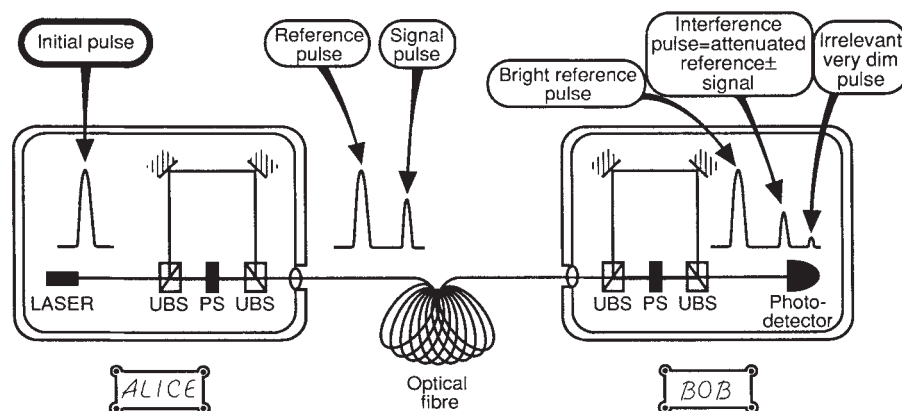
Bennett's proposed apparatus encodes information in the choice between two non-orthogonal states of a quantum system. Some of this information can later be recovered by the legitimate users for use as a key — but the quantum-mechanical uncertainty principle prevents an eavesdropper from listening in without disturbing the states and revealing his activities. Thus the legitimate users are protected against undetected eavesdropping even by an adversary with superior technology.

Although truly unbreakable ciphers, called one-time pads, have been known since about 1917, they all suffer from a serious practical drawback, namely the key distribution problem — the need for would-be users to carry around with them an enormous amount of secret key information, equal in volume to all the messages they might later wish to send. The brilliant mathematical discovery, in the mid-1970s, of public-key cryptography managed to resolve this problem, but unfortunately neither of the proposed public-key cryptosystems has been proved to be totally secure. Instead, it is hoped, without proof, that even the best attack would require an inordinate amount of time. However, this belief may turn out to be false. Indeed, one of the early public-key cryptosystems, the knapsack code, was cracked in 1982.

It is here that physicists now contribute, pointing out that interception is, after all, a measurement performed on a carrier of information and that what can be measured and how depends exclusively on the laws of physics. The original idea of quantum encoding based on quantum uncertainty was put forward by Stephen Wiesner in 1970² and subsequently developed by Bennett and Gilles Brassard who, in 1984, proposed the first feasible quantum key distribution scheme³. Five years later Bennett and

Brassard, with their students, built an apparatus implementing the system⁴. In Europe, slightly different quantum cryptosystems, based on the EPR (Einstein-Podolsky-Rosen) quantum correlations, were analysed by David Deutsch and myself, and the experimental work was undertaken by John Rarity and Paul Tapster from the British Defence Research Agency⁵⁻⁷.

Bennett is now pushing quantum cryptography further towards practicality.



Alice and Bob are in possession of identical Mach-Zehnder interferometers with phase shifters (PS) in the short arms and unsymmetric beam splitters (UBS) such that incoming pulses are split into dim pulses which travel in the short arms and bright pulses which are delayed in the long arms. Phase shifters can be switched in each interferometer between two different positions to permit encoding on Alice's side and filtering on Bob's side. If Alice's and Bob's phase shifts are equal, constructive interference between the signal pulse and the attenuated reference pulse will occur in Bob's interferometer, which can result in a registered photocount in Bob's detector. If Alice's and Bob's phase shifts differ, the destructive interference will prevent photons arriving at the detector. The delayed bright pulse at Bob's detector serves to confirm that the reference pulse has actually arrived and excludes the possibilities of various attacks by substitutions and suppressions.

First, he suggests that encoding should be trimmed down to two quantum states only (four states were employed before). And second, he proposes a practical interferometric realization of the system.

The first part of Bennett's paper takes in some of the basic principles of quantum key distribution. Alice and Bob, who traditionally represent legitimate users, want to establish a cryptographic key. Alice cannot simply generate the key by a random process, make a copy of it and send it to Bob. She knows that classical physics allows the eavesdropper to perform a totally passive measurement, to make an exact copy of the carrier of information without disturbing the original, and read the encoded bits. But if Alice and Bob can both exchange classical public messages, which can be passively monitored but not altered or suppressed by the eavesdropper, and use quantum transmission that cannot be passively monitored, then they will be able to distribute the key safely.

Alice can start the key distribution by sending to Bob a random binary sequence encoded on quantum carriers using two non-orthogonal states $|0\rangle$ and $|1\rangle$ to denote bits 0 and 1 respectively. Bob has two filters that select states orthogonal to $|0\rangle$ or orthogonal to $|1\rangle$, which we will call NOT-0 and NOT-1 respectively. Filter NOT-0 will definitely suppress Alice's bit of value 0 but may with a certain probability transmit Alice's bit of value 1; filter NOT-1 will operate likewise. For each incoming carrier of information, Bob decides randomly and independently of Alice whether to use the NOT-0 or NOT-1 filter and then, after the quantum transmission, publicly

tells Alice in which instances his measurement had a positive result, but without saying whether he used NOT-0 or NOT-1 filter. Alice and Bob discard all instances which failed to produce a positive result, so that the remaining instances should be perfectly correlated, so consisting entirely of instances in which Alice sent 0 and Bob measured NOT-1 or Alice sent 1 and Bob measured NOT-0.

This correlation is perfect only if there was no eavesdropping on the line. The eavesdropper does not know in advance about Alice's and Bob's choices, so he is bound to introduce some errors as he cannot build an apparatus that would definitely distinguish between Alice's 0 and 1 and any substitution of a faked signal may give a positive result when it should not. Alice and Bob check for eavesdropping by revealing to each other in public some random sub-sequence of bits which they subsequently discard. If the test is negative, the distribution must

be set up again; if the test is positive the remaining unrevealed bits form the key.

In practice, Bennett suggests using dim coherent light pulses differing in phase relative to an accompanying bright reference pulse. Pairs of such dim pulses are non-orthogonal and can be both produced and detected in Mach-Zehnder interferometers (see figure). This system, unlike some earlier quantum cryptosystems, has one big advantage — namely, it does not employ photon polarization to encode the key. As quantum information cannot be amplified without losing its quantum character, dropping the need to preserve the photon polarization over long distances gives us a chance of providing the basic components for a long-distance key-distributing apparatus. Indeed, with low-

attenuation optical fibres (0.17 dB km^{-1}) and new silicon photodetectors, Bennett's approach should make it possible to distribute keys over several kilometres rather than the present 30 cm, definitely bringing quantum cryptography into the realm of the plausible. □

Artur Ekert is in the Department of Physics and Merton College, University of Oxford, Oxford OX1 4JD, UK.

1. Bennett, C. H. *Phys. Rev. Lett.* **68**, 3121–3124 (1992).
2. Wiesner, S. *Sigact News* **15**, 78–88 (1983).
3. Bennett, C. H. & Brassard, G. *Proc. IEEE Int. Conf. Computers, Systems and Signal Processing, Bangalore*, 175 (1984).
4. Bennett, C. H., Bessette, F., Brassard, G., Salvail, L. & Smolin, J. *J. Crypt.* **5**, 3–28 (1992).
5. Deutsch, D. *Proc. R. Soc. A* **400**, 97–117 (1985).
6. Ekert, A. K. *Phys. Rev. Lett.* **67**, 661–663 (1991).
7. Bennett, C. H., Brassard, G. & Ekert, A. K. *Scient. Am.* (to appear October 1992).

CANCER

p53, guardian of the genome

D. P. Lane

RESEARCH on the p53 protein is at fever pitch. In the past couple of years it has become clear that inactivation of its tumour-suppressor activity is an almost universal step in the development of human cancers¹. But what does normal p53 do that protects us so efficiently from cancer? I believe that we now know the answer, and it is one that has profound implications for treatment of the disease.

One of the holy grails of those in the p53 field has been to find its associates, cellular proteins that bind to it and regulate its function. By microsequencing, Momand *et al.*² have identified a rat protein that co-immunoprecipitates with p53 (ref. 3) as the rat homologue of the mouse MDM2 protein. The *MDM2* gene (ref. 4) was originally identified as a dominant transforming oncogene present on a 'mouse double minute' chromosome. On page 80 of this issue⁵, Oliner *et al.* now describe the isolation of the human homologue of *MDM2* and the mapping of it to the long arm of human chromosome 12, and show that it too can bind p53.

Amplification

The most exciting aspect of their report is that 17 out of 47 human sarcomas show gene amplification of the *MDM2* locus, a finding which is consistent with previous demonstrations of chromosomal aberrations at this site in sarcomas. High-level expression of the MDM2 protein has also been shown in a cell line with *MDM2* gene amplification. Initial results suggest that in those tumours where *MDM2* is amplified no mutations in p53 are found. So high levels of

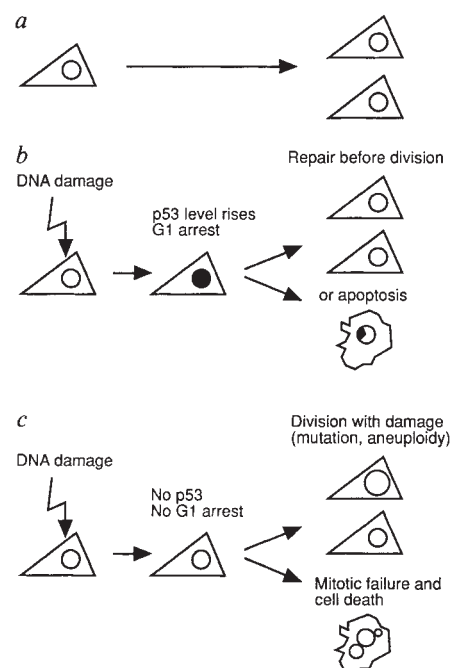
MDM2 may, like the DNA tumour virus oncoproteins, simian virus 40 large T antigen, adenovirus E1b and papilloma virus E6, inactivate the tumour-suppressor activity of p53 by complexing to it. Amplification of MDM2 production may have the same functional effect as mutation of the *p53* gene. The fact that the first associate of p53 to be identified is an important oncogene in its own right will only intensify the search for other partners. It seems likely that they exist, because the different DNA virus oncoproteins that bind p53 occupy different sites on the protein. The initial results with MDM2 suggest that it is not binding to p53 in the same way as SV40 T antigen because, unlike T antigen, it binds to the mutant as well as the wild-type conformation of p53 as defined by monoclonal antibodies³.

What is this function of p53 that is inactivated by mutation or by binding to other proteins? A flurry of papers suggests that the function requires its sequence-specific binding to DNA. The p53 protein was found some time ago to bind to double-stranded DNA and last year it was demonstrated that binding was sequence specific^{6,7}. One consensus binding site has now been closely defined by using the pure p53 protein to select genomic fragments from a random human library and comparing the sequence of the isolated fragments⁸. Strikingly, all the point-mutated p53 proteins analysed have lost this sequence-specific DNA-binding function. Furthermore many of the mutant proteins can act as dominant negatives to inhibit this activity of the wild-type p53.

Classically, this dominant-negative

function is achieved by the formation of inactive hetero-oligomers between the mutant and wild-type p53 proteins; p53 probably binds to DNA as a tetramer. The oligomerization function resides in the carboxy terminus of p53 outside the area where most mutations occur⁹. Consistent with this model is the inhibition of DNA-binding activity of p53 by both SV40 large T antigen⁷ and by MDM2. Of course, because both T antigen and probably MDM2 are DNA-binding proteins and transcription factors in their own right, the hetero complexes may have changed their DNA-binding specificity rather than lost all function.

In addition to this DNA-binding function of p53, the amino-terminal region of the protein, when fused to the DNA-binding domain of the Gal-4 protein, can act as a potent transcriptional activator of genes cloned downstream of a Gal-4 binding site^{10,11}. Interestingly, many point mutations in p53 also abolish this function, from which it seems that they are in some way altering the folding of p53 to hide the activating domain from the transcription machine. Mutant p53 proteins and the adenovirus E1b protein¹² can act dominantly to suppress transcriptional activation in these Gal-4 systems. Given this, the obvious step was to combine the transactivation func-



A model for the function of p53. *a*, Normal cell division, for which p53 is not required. *b*, In the response of a normal cell to DNA damage, the genome-guarding function of p53 is induced. *c*, Cells in which the p53 pathway is inactivated by mutation of p53, or by host (MDM2) or viral oncoproteins, replicate damaged DNA, resulting in mutation, aneuploidy, mitotic failure and cell death. Malignant clones arise from the survivors of this genetic scrambling.

tion of p53 with its sequence-specific DNA-binding function to see if the protein would act as a specific transcription factor in its own right.

Several groups have now taken this step by placing the p53 binding consensus sequence upstream of a minimal promoter and a reporter gene, and then looking at the effect of p53 on the expression of the reporter. The experiments work beautifully *in vivo*, both in yeast and human cells^{13,14}. Only the wild-type protein is active, and again the inhibitory effect of dominant-negative p53 mutants can be clearly seen. It is very satisfying that the temperature-sensitive mutant of p53 that is sensitive for growth suppression is also sensitive for transcriptional transactivation¹³.

Identification

In a notable paper that will allow the rapid identification of the transcription factors with which p53 interacts, Farmer *et al.* (page 83 of this issue¹⁵) now show that this transcriptional activation can readily be measured *in vitro* using purified p53 protein, and that the inhibitory effects of mutant p53 and of SV40 large T antigen can be quantified. So p53 emerges as a potent transcription factor. But why is it a tumour-suppressor gene?

For the answer we must go back to 1984 when it was found¹⁶ that treatment of cells with ultraviolet light or radiomimetic drugs induced the accumulation of normal p53 through a post-translational stabilization mechanism. This dramatic effect is also brought about by γ -irradiation¹⁷ and by a range of cancer chemotherapeutic drugs that damage DNA (M. Fritsche and G. Brandner, personal communication). Kastan and his colleagues¹⁷ have shown that this accumulation of p53 mediates arrest of the cell cycle at G1, which is in accord with the known growth-inhibitory properties of high levels of wild-type p53.

Suddenly everything seems to drop into place. One can propose the following model (see figure), one reminiscent of the bacterial SOS response. Normal p53 acts as a 'molecular policeman' monitoring the integrity of the genome. If DNA is damaged, p53 accumulates and switches off replication to allow extra time for its repair. If the repair fails, p53 may trigger cell suicide by apoptosis¹⁸. Tumour cells in which p53 is inactivated by mutation, or by binding to host or viral proteins, cannot carry out this arrest. They are therefore genetically less stable and will accumulate mutations and chromosomal rearrangements at an increased rate, leading to rapid selection of malignant clones.

This model provides a satisfactory explanation for why DNA viruses must knock out p53 (they must replicate their

DNA within the cell nucleus, but not trigger the damage response as they need the cell to be in S phase for their own replication). It is also consistent with the otherwise puzzling phenotype of p53 null mice¹⁹, which develop normally but have a very high incidence of tumours, and with the high level of aneuploidy and mutational events in tumours. It also explains the genetic instability of fibroblasts from Li-Fraumeni patients²⁰ and, finally, the relative success and risks of conventional radiation and chemotherapy. Tumour cells without working p53 are more susceptible to the killing effect of DNA-damaging agents because they replicate through the damage; similarly, at lower doses, they are also more susceptible to the mutagenic effects of these agents.

We are now in the happy position of having a biochemical function for p53 (as a specific transcription factor) and a biological function (as a G1 checkpoint control for DNA damage). We now need to work out the pathway by which one achieves the other. Following that pathway will almost certainly turn up many hitherto unknown oncogenes and result in modifications to the current use of radiation and chemotherapy in the treatment of cancer. The most appealing option would be to induce p53 with a nontoxic agent to arrest the normal cells

and then treat the tumour with a high dose of the conventional agent, thus radically increasing therapeutic effectiveness. The next few months promise to deliver new highpoints in the p53 story.

D. P. Lane is in the Cancer Research Campaign Laboratories, University of Dundee, Dundee DD1 4HN, UK.

- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. *Science* **253**, 49–53 (1991).
- Momand, J., Zambetti, G. P., Olson, D. C., George, D. L. & Levine, A. J. *Cell* **69**, 1237–1245 (1992).
- Barak, Y. & Oren, M. *EMBO J.* **11**, 2115–2121 (1992).
- Fakhrazadeh, S. S., Trusko, S. P. & George, D. L. *EMBO J.* **10**, 1565–1596 (1991).
- Oliner, J. D., Kinzler, K. W., Meltzer, P. S., George, D. L. & Vogelstein, B. *Nature* **358**, 80–83 (1992).
- Kern, S. *et al.* *Science* **252**, 1708–1711 (1991).
- Bargonetti, J., Friedman, P. N., Kern, S. E., Vogelstein, B. & Prives, C. *Cell* **65**, 1083–1091 (1991).
- El-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. *Nature Genetics* **1**, 45–49 (1992).
- Sturzbecher, H.-W. *Oncogene* (in the press).
- Raycroft, L., Wu, H. Y. & Lozano, G. *Science* **249**, 1049–1051 (1990).
- Fields, S. & Jang, S. J. *Science* **249**, 1046–1049 (1990).
- Berk, A. J. & Yew, P. R. *Nature* **357**, 82–85 (1992).
- Scharer, E. & Iggo, R. *Nucleic Acids Res.* **20**, 1539–1545 (1992).
- Kern, S. E. *et al.* *Science* **256**, 827–830 (1992).
- Farmer, G. *et al.* *Nature* **358**, 83–86 (1992).
- Malzman, W. & Czyzyk, L. *Molec. Cell. Biol.* **4**, 1689–1694 (1984).
- Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. *Cancer Res.* **51**, 6304–6311 (1991).
- Yonish, R. E. *et al.* *Nature* **353**, 345–347 (1991).
- Donehower, L. A. *et al.* *Nature* **356**, 215–221 (1992).
- Bischoff, F. Z. *et al.* *Cancer Res.* **50**, 7979–7984 (1990).

MINERAL DEPOSITS

Continents ring the changes

Robert Kerrich

THE cyclic aggregation and breakup of the continents, the so-called supercontinent cycle, may have controlled the origin of various important metallic mineral deposits. This is the conclusion of Barley and Groves in a remarkable new approach to economic geology¹.

Mineral deposits have not been laid down steadily through Earth history. Rather, certain classes of metal associations appear as extensive metallogenic provinces at specific intervals over geological time, but the reasons for this have been obscure. Examples include abundant mesothermal gold-silver vein deposits developed in the late Archaean, 2,600 million years (Myr) ago, and in the Mesozoic; copper deposits in clastic sediments 2,000, 1,000 and 400 Myr old; and Abitibi-type volcanic-hosted Cu, Zn, Pb deposits 2,700, 1,800 and 400 Myr old. Other classes of metal deposits appear to have a singular principal expression: for instance, Kambalda-type Ni ores were all deposited over 2,000 Myr ago, and Ti-V in anorthosites peaked 1,200 Myr ago (see figure).

Windley², Hutchinson³ and Meyer⁴ in-

dependently recognized peaks in the abundance of different classes of mineral deposit at specific intervals in Earth history, with the corollary of their general absence during intervening times. They formulated the idea that, depending on the deposit type, this secular variation could be attributed to one of three processes: evolution of the oxygen content of the atmosphere-hydrosphere system, decrease of global heat flow, and long-term tectonic trends.

Dissolution and precipitation of some metals are sensitive to oxidation state, and accordingly an increase of the near-surface oxygen content may have controlled the time some metallic ores, including iron formations and unconformity-hosted uranium, were deposited. But there are profound disagreements over the extent and timing of possible changes in oxygen activity, and there are compelling alternative explanations for the temporal distribution of these deposits⁵. A secular decrease of mantle temperature is probably responsible for the restriction of Kambalda-type komatiite-associated nickel deposits to

rocks older than 2,000 Myr, as unusually high temperatures are required to generate magmas of this composition.

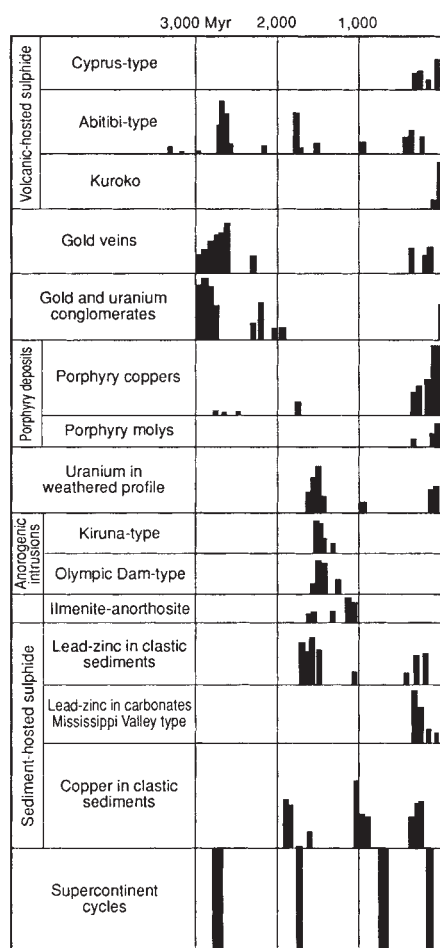
Long-term tectonic trends have been viewed as an evolution of the Earth's crustal development in three stages from an unstable Archaean crust (over 2,600 Myr ago), through the establishment of stable Proterozoic continents (2,600–600 Myr ago), to the operation of plate tectonics in the Phanerozoic (the past 600 Myr). The temporal distribution of mineral deposits has been linked with considerable success to this evolving tectonic pattern. Crustal-evolution models, however, do not explain multiple pulses of certain metallogenic provinces.

In the 1970s, Mitchell and Garson⁶ and Sawkins⁷ developed the first comprehensive syntheses of the relationship between different classes of metal deposits and plate-tectonic geodynamic settings (such as ocean ridges, oceanic and continental magmatic arcs, oceanic and continental hotspots, continent-continent collisions, back-arc basins and continental rifts). For example, they convincingly demonstrated how diamond and tin-uranium deposits formed in association with anorogenic continental hotspots; how deposits of salt- and limestone-hosted Pb–Zn formed at passive margins; how granite-related ores of Cu and Mo and base-metal massive sulphides evolved at magmatic arcs; and how sandstone-hosted U and Cu developed within intracontinental basins. This synthesis worked convincingly for the Phanerozoic, but extrapolations to the Proterozoic and Archaean met with problems stemming from uncertainties in the nature of plate tectonics during these eras. Moreover, this approach does not explain the cyclic nature of some metallogenic provinces now tackled by Barley and Groves.

Extension

The supercontinent cycle is an extension of plate-tectonic theory^{8,9}. The idea emerged in the late 1980s from the recognition that the continental masses do not move randomly over the Earth's surface, but rather assemble and disassemble in a cyclic pattern on a time-scale of 200–500 Myr. All of the present continents formed a single landmass, Pangaea, which broke up about 180 Myr ago, and several researchers have proposed that there were previous supercontinents that broke up and reassembled.

The mechanisms that drive this cycle are not well understood, but a supercontinent would act as a giant thermal blanket, impeding heat dissipation from the mantle, generating melts and mantle upwelling. Collectively these processes could cause the supercontinent to rift apart, subsequently to reassemble above a complementary mantle downwelling.



Timing and scale of mineral deposits, after Barley and Groves¹. The width of each bar represents 50 Myr; the height is the tonnage deposited, as a fraction of the total tonnage for each style over geological history. Lowermost black bars represent approximate age of supercontinent aggregation.

Alternatively, the high-angular momentum of a supercontinent on the spinning Earth may generate stresses that induce self rupture.

Barley and Groves's central contribution is the recognition that cyclicity in the timing of specific classes of metal deposit coincides with episodes of supercontinent breakup and reassembly (see figure). Metal deposits related to anorogenic magmatism (sandstone-hosted Cu and Pb) would be most abundant during initiation of supercontinent fragmentation, whereas deposits related to convergent tectonics (porphyry Cu, and volcanic-hosted Cu, Pb, Zn) predominate during periods of subduction and supercontinent aggregation. Thermal decay, reflected in evolution of the lithosphere, is responsible for the singular expressions of other metallogenic provinces through the Earth's history.

The scarcity of porphyry-Cu deposits in the Precambrian, and change in style of volcanic-hosted Cu, Pb, Zn deposits from Archaean Abitibi type, through

Proterozoic type, to the Phanerozoic Kuroko and Cyprus types, may reflect differences in style of subduction and arc magmas, that in turn stem from decreasing thermal gradients.

Aspect

Another important aspect of thermal decay is that Archaean crust is bonded to a stable, refractory mantle lithosphere keel about 200 km thick, which promotes the preservation of Archaean crust. Post-Archaean mantle temperatures were too low to generate such a thick refractory keel. This distinction in preservation potential is not generally incorporated into studies of the temporal distribution of metal deposits.

Superimposed on this cycle and thermal trend are variations arising from preservation and subtleties of tectonic style. The scarcity of porphyry Cu and epithermal Au ores in rocks older than 200 Myr is widely considered to be the consequence of their low preservation potential in rapidly eroded magmatic arcs and collisional mountain belts.

Barley and Groves argue that preservation potential is higher in peripheral (Cordilleran style) than internal (Himalayan style) mountain belts, but some may disagree. And the tendency for mesothermal Au–Ag vein deposits of all ages to be largest and most abundant at oblique convergent margins¹⁰, rather than at internal orogens, may result from subtleties of this specific tectonic regime. For example, the Californian Mother Lode deposits and Archaean gold–silver vein deposits are related to large fluid volumes mobilized through regional faults in these circumstances.

The combination of the supercontinent cycle and decreasing heat flow, reflected in the evolution of the lithosphere and tectonic style, gives economic geologists a remarkable framework for understanding the timing and distribution of many of the mineral deposits we have come to rely on. □

Robert Kerrich is in the Department of Geological Sciences, University of Saskatchewan, Saskatoon, Canada S7N 0W0.

1. Barley, M. E. & Groves, D. I. *Geology* **20**, 291–294 (1992).
2. Windley, B. F. *The Continents* (Wiley, Chichester, 1974).
3. Hutchinson, R. W. in *Evolution of the Earth* (eds O'Connell, R. J. & Fyfe, W. S.) 120–140 (Am. Geophys. Union, Washington, DC, 1981).
4. Meyer, C. A. *Rev. Earth planet. Sci.* **16**, 147–172 (1988).
5. Clemens, H. & Badham, N. *Geology* **10**, 141 (1982).
6. Mitchell, A. H. G. & Garson, M. S. *Mineral Deposits and Global Tectonic Settings* (Academic, London, 1981).
7. Sawkins, F. J. *Metal Deposits in Relation to Plate Tectonics* 2nd edn (Springer, Berlin, 1990).
8. Murphy, J. B. & Nance, R. D. *Scient. Am.* **84**–91 (April 1992).
9. Hoffman, P. E. A. *Rev. Earth planet. Sci.* **16**, 543–604 (1988).
10. Wyman, D. A. & Kerrich, R. *Econ. Geol.* **83**, 451–456 (1988).

On to molecular mechanisms

Charles F. Stevens

FOR the past several years, molecular neurobiologists have been fascinated by a series of worldwide races to clone the glutamate receptors found at the brain's most common synapses. The initial contest was won several years ago by researchers in Heinemann's laboratory¹, who identified the first member of this large gene family. Another notable victory occurred eight months ago when Nakanishi's team found² the first member of the NMDA receptor subfamily, a special relative of the first glutamate receptor. Nakanishi's report instantly set off another race to find the anticipated NMDA receptor relatives. Now two groups, Sakmann and Seeburg in Germany³ and Mishina *et al.*^{4,5} in Japan, have succeeded in finding three other members of the NMDA receptor subfamily. The second of the papers from Mishina's group appears on page 36 of this issue.

Why should anyone care about finding a few more members of a family that already consists, published and unpublished, of well over a dozen subunits? The reasons have to do with the importance of the NMDA receptors and the richness of their properties.

Although glutamate is the actual neurotransmitter for the entire family, glutamate analogues define three pharmacologically distinct receptor groups present at most of the brain's excitatory synapses — the AMPA, kainate and NMDA receptors. This pharmacological classification seems to make evolutionary sense in that branches of the family tree, based on sequence similarity, can be identified for each pharmacological type. Of the three types, the NMDA receptor has received the most attention in the past several years because it plays a central part in so many interesting brain functions and dysfunctions.

First, NMDA receptors provide the trigger event for producing long-term potentiation, the persistent synaptic modification that most neuroscientists believe is the basis for a form of memory. Understanding the NMDA receptor therefore offers the exciting possibility of explaining one of the brain's essential cognitive functions right down to the detailed molecular level. Second, NMDA receptors are intimately involved with the phenomenon of excitotoxicity: in certain pathological situations, such as stroke and head trauma, much of the damage occurs because the nerve cells become excessively active and literally excite themselves to death. Here the possibility arises that altering NMDA receptor function with appropri-

ate drugs could prevent excitotoxicity and thereby save many neurons that otherwise would die.

Third, some drugs that block NMDA receptor function — for example, phencyclidine, known on the street as 'angel dust' — cause bizarre psychosis-like symptoms in humans. Perhaps, then, interference with NMDA receptor function caused in other ways, maybe even through mutations, contributes to some psychoses. Sufficient knowledge about the NMDA receptor might then provide a rational basis for treating such disorders.

Fourth, NMDA receptors have been implicated in the rewiring of the brain that occurs, during brain development, with experience. Because anatomically different brains result from different experiences (experience here can range from seeing things to having epileptic seizures), understanding this vital process could be of great practical as well as intellectual importance.

Finally, NMDA receptors are involved in some of the computations carried out by the brain. To give a simple but possibly important example, they are used for synaptic transmission in pain pathways, and special NMDA receptor properties, and modifications of these properties, are responsible for a hypersensitivity exhibited by injured regions⁶. Understanding the mechanisms for this process already suggests practical measures for reducing post-operative pain.

What, though, are some of the NMDA receptor properties that help account for these various roles?

The postsynaptic proteins responsible for synaptic transmission are used to detect the presence of neurotransmitters released presynaptically: glutamate, for example, binds to glutamate receptors, produces a conformational change that leads to the flow of ionic current and thus to the voltage changes used by a neuron for its communications and computations. These postsynaptic receptor proteins must not be — and are not — influenced by the voltage changes they and other ion channels produce, for if they were the activity of one synapse could spoil the effect of another. The neurotransmitter receptors are thus 'ligand gated', that is, they are only operated by the presence of their ligand and not by voltage. This is in contradistinction to the voltage-gated channels, the ones that sense only the neuron's voltage and are responsible for nerve impulses.

The exception to this ligand-gated/

voltage-gated dichotomy is the NMDA receptor, which requires both the presence of glutamate (released by a presynaptic axon) and an appropriate change in the neuron's voltage in order to operate. This conjoint ligand and voltage sensitivity lies at the heart of many NMDA receptor functions, from detecting the correlated neuronal activity necessary for long-term potentiation and learning, to giving pain its characteristic properties.

Another interesting, and quite possibly unique, characteristic of NMDA receptor channels is their requirement for glycine as a co-agonist. Although called 'glutamate' receptors, actually they refuse to work unless glycine is also present. This glycine binding seems important, although we have no clear idea of its functional significance.

In addition to the dual agonist/voltage dependence and coagonist activation, NMDA receptors have a host of other characteristics, ranging from ionic selectivity through especially complex kinetics of operation, to regulation by second-messenger systems.

What the two laboratories have discovered, then, is the existence of three new subunits that join with Nakanishi's original NMDAR1 protein to form the brain's NMDA receptors. These additional subunits are inconveniently large, close to 150,000 in relative molecular mass, and are not very closely related to other members of the glutamate receptor family (having only about 20 per cent sequence similarity); they are clearly family members, though, because certain regions, especially those that span the membrane, are quite well conserved. Although Nakanishi's NMDAR1 is everywhere, the newly discovered subunits have a different distribution in the brain, with one type concentrated, for example, in the cerebellum and another in the forebrain.

Moreover, the different subunit combinations (NMDAR1 plus one of the new ones) have distinctly different functional characteristics in expression systems. For example, they differ in the degree to which voltage as well as ligand binding is required for operation, and the extent to which the glycine co-agonist is required, and in their sensitivity to pharmacological blockade.

With the identification of these additional NMDA receptor subunits it now becomes possible to understand the molecular mechanisms responsible for the NMDA receptor's rich repertoire of characteristics. Firm indications about the origin of, say, ion-channel selectivity, will come at once from comparing the characteristics of receptors with different subunit combinations, and these clues will then form the basis of structure/function studies using site-

directed mutagenesis. Moreover, having the array of subunits will permit pharmacologists to develop drugs that modify the operation of specific subunit combinations, and this is the key both for unravelling the role of NMDA receptors in normal brain function and in disease, and also as a basis for developing useful drugs.

These races have gripped the attention of the neurobiological community. Now the longer process of using the clones to work out mechanisms, a process already

well under way for the non-NMDA receptors, can begin for the NMDA receptor subfamily. □

Charles F. Stevens is at the Salk Institute, La Jolla, California 92037, USA.

1. Holmann, M., O'Shea-Greenfield, A., Rogers, S. W. & Heinemann, S. *Nature* **342**, 643–648 (1989).
2. Moriyoshi, K. et al. *Nature* **354**, 31–37 (1991).
3. Monyer, H. et al. *Science* **256**, 1217–1221 (1992).
4. Meguro, H. et al. *Nature* **357**, 70–74 (1992).
5. Kutsuwada, T. et al. *Nature* **358**, 36–41 (1992).
6. Chen, L. & Huang, L.-Y. M. *Nature* **356**, 521–523 (1992).

ARCHAEOLOGY

Under the skin of Nazca

Warwick Bray

WHY the Nazca lines were constructed remains a matter of conjecture^{1,2}. But the age of the lines has now been established more precisely by the application of a new technique^{3,4}, the upshot of a collaboration between an archaeologist (Persis Clarkson) and a geomorphologist (Ronald Dorn).

The Nazca lines, of south coastal Peru, are one of the world's largest and most spectacular archaeological monuments. Hundreds of square kilometres of the desert surface are covered with a palimpsest of immense geoglyphs, or ground-drawings, made by sweeping aside the dark, patinated stones to expose the lighter soil beneath. Designs include animals (quadrupeds, birds, spiders, killer whales) as well as spirals, triangles and trapezoids, spoke-like arrangements and straight lines several kilometres long.

Previous attempts to date the lines have given inconclusive results. It has long been realized that the animal designs can be matched with the painted motifs on Nazca pottery dating to about 200 BC – AD 600. But the linear and geometric ground-drawings offer no such stylistic clues and, as surface features, they have no stratigraphic context. Remains of wooden posts at the intersections of two different sets of lines have given (uncalibrated) radiocarbon dates of AD 525 ± 80 and 490 ± 80, but attempts to date the geoglyphs by means of the ancient pot sherds found on the cleared surfaces have produced contradictory information. Two investigations concluded that most of the pottery, though not all of it, belonged between 200 BC and AD 400; a third found evidence that the trapezoids, at least, may be much more recent (AD 600–1476)⁵.

All of these approaches are indirect and unsatisfactory. The posts could have

been reused or reset; the pottery fragments, assuming they have not been disturbed by later visitors, provide limiting dates but not the actual date of construction. What was needed was a method for dating the lines themselves. This is what Dorn's team has provided, by introducing a technique for determining the age of the rock varnish on stones disturbed by the building of the lines.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Condor-like figure on Pampa de San José, Peru.

Rock varnish is a dark coating, made up of clay minerals and manganese and iron oxides, that builds up on stone in dryland environments. Electron microscopy revealed that organic matter (lichens, cyanobacteria, fungi) adhering to the surfaces of the rock may become trapped beneath the varnish accretion, which begins to form whenever a fresh surface is exposed to the elements. Organic detritus of this kind sometimes occurs in quantities sufficient for accelerator mass spectrometry radiocarbon dating. Nine samples were taken from stones removed by the line-builders. After correction for tree-ring calibration, the carbon-14 dates ranged from about 190 BC to AD 600, an excellent fit with the 'traditional' dating for the geoglyphs. These carbon-14 determinations are minimum ages, however,

because the organic matter must be older (though probably not much older) than the varnish which covers it.

The same technique was applied to another long-standing problem of Nazca archaeology, the date of the filtration-galleries, or *puquios*. These are underground tunnels, analogous to the qanats of the Near East, which were cut into hillsides to tap subterranean aquifers for irrigation purposes. In the Old World qanats first appear in Assyrian times (8th century BC) in Iraq, and spread to the Mediterranean with the Roman and Arab conquests. There is sound historical evidence to show that the Spanish conquistadors transferred the technology to both Peru and Mexico⁶. The mass of documentary information from colonial Peru contradicts the popular belief that the filtration-galleries of the Nazca region are of pre-Spanish age and were linked with a group of settlements contemporary with the ground-drawings. At this stage, discussion had reached a stalemate — none of the colonial documents refers specifically to the Nazca galleries, and the archaeological case is at best circumstantial.

The matter has been partially resolved with the dating by Dorn and colleagues of the varnish on stone lintels from two of the Nazca *puquios*. They obtained calibrated ages of AD 552–644 and 591–698 (error margin 1 sigma). If these figures are confirmed with further samples, some of the Peruvian galleries, through clearly not all of them, must be precolumbian, and Peruvian Indians must be credited with an early and independent invention of qanat technology. For reasons unknown, the filtration-galleries of Nazca (like the drained fields of the Andes⁷) had already fallen into disuse by the time of the Inca expansion, and are not mentioned by the first European chronicles. As so often happens, the archaeologists seem to have found the answer to one question but created another. □

Warwick Bray is at the Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK.

1. Morrison, T. *Pathways to the Gods: The Mystery of the Andes Lines* (Michael Russell, Salisbury, 1978).
2. Aveni, A. F. (ed.) *The Lines of Nazca* Mem. Am. Phil. Soc. Vol. 183 (American Philosophical Society, Philadelphia, 1990).
3. Clarkson, P. B. & Dorn, R. I. *Boletín de Lima* **13**, 33–45 (1991).
4. Dorn, R. I., Clarkson, P. B., Nobbs, M. F., Loendorf, L. L. & Whitley, D. S. *Ann. Ass. Am. Geogr.* **82**, 136–151 (1992).
5. Silverman, H. & Browne, D. *Antiquity* **65**, 208–220 (1991).
6. Barnes, M. & Fleming, D. *Latin American Antiquity* **2**, 48–68 (1991).
7. Bray, W. *Nature* **345**, 385 (1990).

Tony Morrison/South American Pictures

Melting of wet lithosphere

Richard W. Carlson

SCATTERED about the continents, eight distinct regions have experienced floods of basaltic volcanism in the past 250 million years. These flood basalt provinces are marked by eruptions of as much as a million cubic kilometres of lava over just a couple of million years or less. Trace-element and isotope compositions of many flood basalts are reminiscent of the continental crust, yet the crust and the section of rigid mantle beneath the continents, together known as the lithosphere, generally are assumed to be too cold to allow the high melting rates implied by flood volcanism. What then is the source of the magma? Elsewhere in this issue (*Nature* 358, 57–59; 1992), K. Gallagher and C. Hawkesworth suggest that the answer may lie in the abundance of water in the lithospheric mantle.

The argument against a lithospheric source for flood basalts, as presented by D. McKenzie and M. J. Bickle (*J. Petrol.* 29, 625–679; 1988) and most recently by N. T. Arndt and U. Christensen in the latest issue of the *Journal of Geophysical Research* (97, 10,967–10,981; 1992), is that the mantle lithosphere beneath continents is believed to be about 300–400 °C below its melting point at a given depth. For the lithosphere to melt, heat must be introduced by conduction from a much hotter portion of the mantle; or the deep lithosphere must rise rapidly to shallower depth so that decompression lowers the melting temperature of the rock more quickly than heat loss cools it. This kind of ascent of deep lithosphere will occur when the overlying continent thins through extension associated with plate-tectonic forces, and most flood basalt provinces are associated with extensional rifts. Neither ascent nor conduction, however, seem to be capable of produc-

ing, from a dry mantle lithosphere, the large quantities of melt seen in flood basalt provinces. Arndt and Christensen's new calculations suggest that the lithosphere provides no more than 2 per cent of the melt produced during the process of rifting a continent to form an ocean basin. Most of the melt comes from deeper, hotter mantle.

This result, in part, has led to increasing acceptance of the idea that flood basalts are caused by the arrival at the Earth's surface of a plume of material from the deep mantle (M. A. Richards, R. Duncan & V. E. Courtillot *Science* 246, 103–107; 1989; I. H. Campbell & R. W. Griffiths *Earth planet. Sci. Lett.* 99, 79–93; 1990). Indeed, many flood basalt provinces mark the starting point of the tracks of volcanic ridges left by mantle hotspots as the surface plates move over them (see figure). A plume origin for flood basalts can account for the high magma production rates observed in flood basalt provinces, but almost all of the melt (99.6 per cent according to Arndt and Christensen) will be derived from the hot plume instead of the shallow lithospheric mantle.

In support of the idea that plume-supplied mantle is involved in flood basalt genesis, some flood basalts have chemical and isotope compositions similar to those of lavas now being erupted at the current location of the hotspot associated with the flood basalt province. For example, many of the 65-million-year-old Deccan flood basalts of India overlap in composition with basalts currently being erupted on the associated hotspot, the ocean island of Réunion. Other flood basalts, however, have trace-element and radiogenic isotope characteristics that show more similarity to rocks of the continental crust than to any volcanic system in the ocean basins.

This is particularly true of the flood basalts associated with the breakup of the southern supercontinent, Gondwana, specifically in the basalts of the Karoo (southern Africa), Paraná (South America), Tasmanian and Transantarctic provinces. Many of the basalts in these provinces are enriched in silicon and a number of trace elements, the so-called incompatible elements, that preferentially partition into melts rather than the solid residues of melting. These basalts also have radiogenic isotope compositions indicating that their incompatible-element enrichment is an ancient feature. This suggests a source material that has remained separate from the convectively stirred mantle for quite some time. A good candidate for this material is ancient continental lithosphere.

How does one then reconcile the chemical signature of lithospheric origin with the physically reasonable models that indicate that lithosphere cannot melt to a sufficient degree to furnish these magmas? One way is Gallagher and Hawkesworth's. The addition of water to the mantle, even in concentrations as low as 0.4 per cent by weight, lowers the melting temperature of the mantle rocks by several hundred degrees. Using similar thermal models to those of Arndt and Christensen, but with a 'wet' lithospheric mantle, Gallagher and Hawkesworth show that large volumes of melt can be produced from the lithospheric mantle. As extension of the lithosphere proceeds, the relative proportion of lithospheric melt that reaches the surface gradually decreases as more of the deeper, sublithospheric mantle melts as it rises to shallower depths. Several flood basalt provinces indeed show a transition in the composition of basalts erupted. The sequence starts with lavas that have strong lithospheric signatures and then tends towards compositions more similar to basalts found in the ocean basins where the lithosphere is young and thin and, hence, cannot contribute to magma genesis. ►



Flood basalt provinces on Earth and their connection with hotspot tracks. CRB stands for Columbia River Basalts. The hotspot associated with this province now occurs at Yellowstone (YS). (From R. W. Carlson *Aust. J. Earth Sci.* 38, 525–544; 1991.)

The observed chemical transition in flood basalts agrees with the order found in Gallagher and Hawkesworth's model, and is contrary to that expected for dry melting models. Under dry conditions, the high temperatures needed for melting require either high degrees of extension, or long times to conduct enough heat into the lithosphere. Both condi-

tions lead to lithospheric melting late in the history of volcanism rather than early. □

Richard W. Carlson is in the Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Road N.W., Washington, DC 20015, USA.

IMMUNOLOGY

A sense of something missing

David H. Raulet

NATURAL killer (NK) cells lyse various tumour cells. That has been known for many years, but why some tumour cells are more sensitive than others to NK lysis was poorly understood, as was the nature of the recognition molecules concerned. Then it emerged that sensitivity to NK cells is inversely correlated with expression of major histocompatibility complex (MHC) class I molecules on target cells. Do class I molecules on the target cells inhibit NK cells? Or do they mask target-cell antigens recognized by NK cells? Are MHC-bound peptides involved? And what is the nature of the NK receptor?

The answers are not in yet. But results reported by Karlhofer *et al.* on page 66 of this issue¹ support a model in which lysis of the target cell is inhibited when an NK receptor, which they identify as Ly-49, is engaged by class I molecules on the target cell. Moreover, a report by Storkus *et al.*² in *Proceedings of the National Academy of Sciences* implicates the peptide-binding groove of class I molecules in recognition by NK cells.

The story has several threads, but begins 30 years ago with the finding that hybrid mice heterozygous at the MHC gene complex, say A × B, can reject bone marrow grafts from a homozygous parental donor, say A (for review see ref. 3). Operationally, the *absence* of a self MHC-linked gene seems to target cells for rejection, contrary to the 'laws of transplantation' that govern the behaviour of grafts rejected by T cells. It took another 20 years to discover that rejection in this instance is mediated by NK cells, not T cells. That the MHC genes responsible encode class I molecules was directly demonstrated in one case⁴, with the use of mice transgenic for an allogeneic class I MHC molecule, D^d. (However, there are indications that non-class-I genes encoded in the MHC also play a part³.)

Meantime, *in vitro* studies of tumour-cell sensitivity to NK lysis began to reveal an inverse correlation with class I MHC expression. The effect of class I was conclusively demonstrated by show-

ing that restoration of class I expression (by transfection) rendered class-I-negative cell lines resistant to NK cells⁵. More recently, it has been shown that even nontransformed class-I-deficient cells, such as bone marrow or lymphoblasts from class-I-deficient mice, are sensitive to NK lysis⁶.

It was then discovered that cloned human NK cells can lyse allogeneic non-transformed lymphoblasts *in vitro*⁷. The sensitivity to lysis of target cells maps to the MHC and is recessive; that is, cells lacking a specific MHC allele are susceptible.

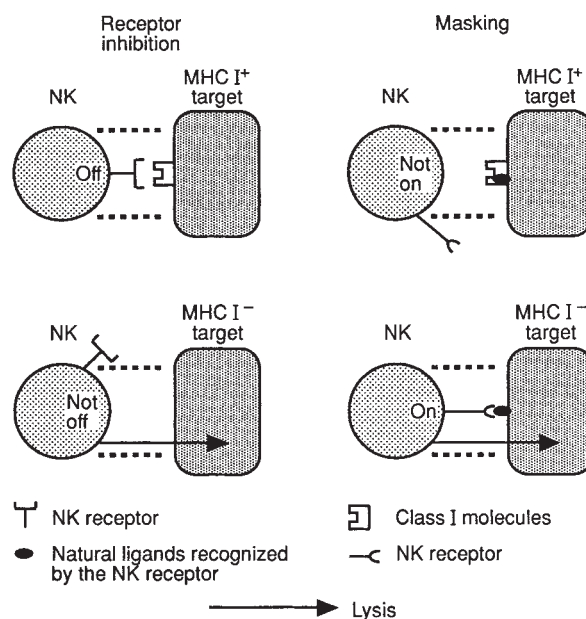
The convergence of all of this work led to the missing-self hypothesis⁸, wherein NK cells destroy autologous cells that have lost or altered expression of self-MHC molecules. How can NK cells perceive the absence of MHC molecules? According to one model (see figure), unidentified antigens recognized by NK cells can be masked by class I molecules. In another, NK cells have receptors for class I molecules, but engagement of the receptor inhibits the cells' lytic activity.

To account for the differential effects of MHC alleles on NK lysis, some diversity (perhaps limited) of NK receptors is suggested in both models. Proponents of the masking hypothesis argue that there is a family of target antigens, individual members of which are masked by some class I antigens better than others; adherents of the receptor-inhibition model argue that the NK receptors are specific for different class I molecules, although there may be a good deal of cross-reaction. And because the

second model proposes that lysis occurs when target cells *fail* to be recognized by the class-I-specific receptor, it must be considered that other receptor-ligand pairs mediate NK-target cell interactions leading to cell lysis. Happily, there is evidence for a plethora of such interactions⁹.

The new report by Karlhofer *et al.* stems from the earlier studies¹⁰ of the Ly-49 cell-surface molecule, which is expressed by about 20% of NK cells in some mouse strains. Believing that Ly-49 might represent a specific NK receptor, Karlhofer *et al.* tested Ly-49⁺ and Ly-49⁻ NK cells against a large panel of tumour target cells¹. Indeed, they found a striking, if imperfect, correlation between MHC genotype and sensitivity of target cells to 'Ly-49⁺ NK cells. H-2^b cells are sensitive, whereas H-2^d and H-2^k cells are resistant. Transfection of H-2^b cells with one of the H-2^d class I genes, D^d, makes the cells resistant, but transfection with K^d or L^d has no effect. In contrast, Ly-49⁻ NK cells lyse all of these target cells.

The correlation of MHC specificity with Ly-49 expression obviously suggests that Ly-49 is involved in specific NK recognition. The earlier work showed that Ly-49 is a member of a family of linked genes, called the NK complex, which encode type II membrane proteins containing an extracellular lectin-like



Two models to account for lysis by natural killer (NK) cells of cells without class I MHC molecules⁴. The receptor-inhibition model implies that the interaction of the NK receptor with class I molecules serves to overrule or inhibit lysis activated by distinct NK-target cell interactions (dashed lines). Karlhofer *et al.*¹ favour a scheme of this sort in which the NK receptor is the Ly-49 molecule and the class I molecule is D^d. The masking model suggests that class I obscures the unidentified natural ligands recognized by the NK receptor. Reaction of the NK receptor with the exposed ligand activates lysis of the target cell.

domain¹⁰. Other members of the family are also implicated in NK function, and an attractive idea is that a subset of these genes, including Ly-49, encode isoforms of the NK receptor, each of which is expressed on a subset of NK cells. In this regard, others have identified distinct mouse and human NK cell-surface molecules which define NK cell subsets with specificity for MHC-linked molecules^{11,12}. It will be of considerable interest to see whether these molecules are also encoded by genes in the NK gene complex.

Do the present studies help distinguish between the two models of NK recognition? Perhaps so, because Karlhofer *et al.* show that lysis of *D^d*-transfected target cells by Ly-49⁺ NK cells can occur if antibodies against either Ly-49 or *D^d* are added to the reactions. This is most simply explained by recourse to the receptor-inhibition model: both antibodies can block delivery of the inhibitory signal, through Ly-49, to the NK cell.

The authors marshal another argument in favour of the receptor-inhibition model, by pointing out that *D^d* expression by target cells even inhibits NK lysis mediated by antibody (antibody-dependent cellular cytotoxicity). The idea is that class I recognition has a veto, and can overrule any signals that might activate NK lysis. But the case for receptor inhibition is not conclusive. The masking model is still possible, because Ly-49-specific antibodies may be triggering lysis rather than blocking inhibition of it; and antibodies against a class I molecule may cause the disassociation of a bound NK antigen. Among other things, proof of the proposition will require evidence that Ly-49 actually binds to *D^d*.

At first glance the results of Karlhofer and colleagues seem concordant with the missing-self hypothesis, but one aspect of the observations is startlingly at odds with it. Their experiments employed NK cells from H-2^b mice, yet these NK cells lyse H-2^b tumour cells, and such lysis is inhibited by introduction of a foreign (*D^d*) H-2 gene. Therefore the results literally favour a 'missing-foreign', rather than a missing-self interpretation. Such a system would make very little sense physiologically — in fact you would expect the NK cells to cause a severe autoreaction.

Two rationalizations of the results come to mind. One is that the Ly-49⁺ NK cells resident in H-2^b mice are normally in an inactive state, but culture of the cells in high levels of interleukin-2, as performed here, results in activation of the cells. The other, favoured by Karlhofer *et al.*, is that Ly-49⁺ NK cells react not only with *D^d*, but also with H-2^b class I molecules to which specific

self peptide(s) is bound. The peptide may be lost from tumour cells, explaining the sensitivity of H-2^b tumour cells to Ly-49⁺ NK cells. In fact, the authors comment that normal H-2^b cells are resistant to Ly-49⁺ NK cells.

Is there reason to believe that specific class-I-associated peptides are involved in NK recognition? No direct evidence on the subject has been reported, but the provocative results of Storkus *et al.*² point strongly to this possibility. Starting with a human class-I-negative tumour cell that is sensitive to NK cells, they found that expression of most class I genes would make the cells revert to NK resistance. One human class I gene, *HLA-A2*, failed to do so. Remarkably, a mutant *HLA-A2*, with a single amino-acid substitution, did cause the cells to revert. This substitution, of histidine to aspartate at position 74 in the $\alpha 1$ domain of *HLA-A2*, is thought from other studies to unblock a side pocket in the peptide-binding groove.

So it seems that the groove is in fact involved, directly or indirectly, in NK recognition. An enticing (if speculative) hypothesis is that NK cells are inhibited by recognition of autologously encoded ligands bound in the class I peptide-binding groove. Loss of the class I molecule from the cell surface would therefore render cells NK sensitive. But loss of the class-I-bound ligand would have the same effect. This might occur because a virus has switched off cellular protein biosynthesis, or because a viral peptide displaces the NK ligand from the class I groove. In this way, NK cells may contribute to host defence against intracellular pathogens. Whatever the role of class I in NK cell activity, it should be noted that NK cells apparently mediate cell lysis by several mechanisms. Some of them may be independent of regulation by class I molecules. □

David H. Raulet is in the Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA.

- Karlhofer, F. M., Ribaud, R. K. & Yokoyama, W. M. *Nature* **358**, 66–70 (1992).
- Storkus, W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 5989–5992 (1991).
- Yu, Y.Y.L., Kumar, V. & Bennett, M. A. *Rev. Immun.* **10**, 189–213 (1992).
- Ohlen, C. *et al.* *Science* **246**, 666–668 (1989).
- Storkus, W. J., Alexander, J., Payne, J. A., Dawson, J. R. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2361–2364 (1989).
- Liao, N., Bix, M., Zijlstra, M., Jaenisch, R. & Raulet, D. *Science* **253**, 199–202 (1991).
- Cicone, E. *et al.* *J. exp. Med.* **172**, 47–52 (1990).
- Kärre, K., Ljunggren, H. G., Piöntek, G. & Kiessling, R. *Nature* **319**, 675–678 (1986).
- Storkus, W. J. & Dawson, J. R. *Crit. Rev. Immun.* **10**, 393–416 (1991).
- Yokoyama, W. M., Jacobs, L. B., Kanagawa, O., Shevach, E. M. & Cohen, D. I. *J. Immun.* **143**, 1379–1386 (1989).
- Sentman, C. L., Hackett, J., Jr., Kumar, V. & Bennett, M. J. *Exp. Med.* **170**, 191–202 (1989).
- Moretta, A. *et al.* *J. exp. Med.* **172**, 1589–1598 (1990).

Terse comments

THE modern definition of a random number or sequence is one which cannot be generated by any program shorter than itself. It follows that any non-random sequence, such as *Nature*, can be generated by a program shorter than itself. Useful economies could therefore be made by publishing the program rather than the journal.

The editor is looking on uneasily as Daedalus tries this out in practice. His approach is to compile a dictionary containing all the words in a given issue. He then replaces the most common word by the shortest binary code, 0, the next common by binary 1, the next by binary 10, and so on. Printed as ASCII character code, this reduces the journal to about 10 per cent of its original length. A second round of compression takes the process further. This looks for common sequences of words, from simple pairs up to quite complex clichés and 'themes with variations', replaces them by binary codes, and enters these in the dictionary as well. A third round of compression packs the code tighter still.

The first few substitutions shrink the journal rapidly. As the process continues, each new entry in the dictionary shrinks the text less and less. Ultimately, a substitution is reached which expands the dictionary just as much as it shrinks the text. At this point the journal has reached its smallest possible coding. Some DREADCO mathematicians favour the converse approach, which starts by looking for large repeated or nearly repeated chunks of text, and only later descends to word level. The final outcome should be the same: a compressed publication consisting of a program (the dictionary) and data (the sequence it operates on), together occupying only a few per cent of the present page space.

Each issue has a slightly different ideal dictionary and code sequence, but the former varies much less than the latter. Further savings could be made by standardizing a globally optimal *Nature* dictionary, and publishing it once only in program form. Each individual issue would then consist of the code sequence alone. A character-reader feeding a computer loaded with the program could reconstruct it.

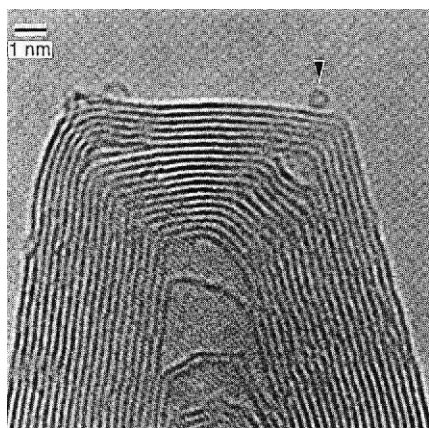
Where *Nature* has gone, other scientific publications will surely follow. Librarians will rejoice as the torrent of paper flooding endlessly towards them dries up to a cryptically printed trickle. Even editors will find the new technology useful. A paper which collapses under coding to a mere few lines of text cannot have very much to say. Conversely, one which refuses to collapse at all is clearly completely random already, and is not worth printing.

David Jones

Smallest carbon nanotube

SIR — The discovery and synthesis of C_{60} molecules have led to an explosion of interest in the family of carbon structures and compounds. We have recently reported the discovery of carbon nanotubes¹, consisting of concentric sheets of carbon hexagons arranged in a helical manner, but with the ends capped by pentagons². The diameters of the tubes are in the nanometre range, with the smallest tube observed so far having a diameter of 2.2 nm. The structures consist of two to many layers. Recently, there has been speculation about the smallest possible carbon fibre whose diameter would correspond to the C_{60} molecule³, and which would have extremely interesting physical properties.

While observing the structure of car-



bon tubes in the transmission electron microscope, we have come across images that could indeed correspond to an isolated tube with the C_{60} diameter (see figure). The spherical hollow structure (arrow) has a shell with an outer diameter of approximately 0.8 nm, which could correspond to an image from a single C_{60} structure⁴. We do not think the structure is a single isolated C_{60} molecule because its contrast is comparable to that obtained from the {002} lattice fringes of the supporting polygonized carbon-tube structure, which is a large tube with many layers of carbon hexagons and hence large numbers of carbon atoms in a diffracting column compared with a single C_{60} molecule. So, we believe that what we are seeing is a tube with the diameter of a C_{60} molecule and the dimension of a few such units in the direction perpendicular to the plane of the image, being observed end-on. The observation of such spherical and elongated structures on the surface of tubes is fairly common, but this is the first time a single isolated structure has been found.

Such extremely small tubular structures could act as nucleation sites for the

growing tubes. Raft-like graphite layers are often seen in the form of incomplete layers on the surface of the tubular assemblies. In the very early stages of nucleation and growth, these rafts of carbon hexagons nucleated on the surface of graphitic sheets of the tube may curl up to form cylindrical shapes of small diameter, like the tube observed in the figure. It is, however, not possible to tell how actual growth proceeds on these structures to produce the completed layers or whether these structures are closed or open at their tube ends.

P. M. Ajayan

Sumlo Iijima

Fundamental Research Laboratory,

NEC Corporation,

34 Miyukigaoka, Tsukuba,

Ibaraki 305, Japan

1. Iijima, S. *Nature* **354**, 56 (1991).
2. Iijima, S., Ichihashi, T. & Ando, Y. *Nature* **356**, 776–778 (1992).
3. Dresselhaus, M. S., Dresselhaus, G. & Saito, R. *Phys. Rev.* **B45**, 6234 (1992).
4. Cox, D. M. *et al. J. Am. chem. Soc.* **113**, 2940 (1991).

Blood substitutes and infection

SIR — Increasing concern about blood-transmitted viral pathogens has spurred the development of a blood substitute that can effectively replace the oxygen-carrying capabilities of red blood cells. Chemically modified haemoglobins and recombinant human haemoglobin synthesized in *Escherichia coli*, *Saccharomyces cerevisiae* and in transgenic animals have been investigated as potential oxygen-carrying red-cell substitutes¹. They turn out not to be suitable because the oxygen affinity is too high to permit efficient tissue oxygenation, and the haemoglobin molecules are not stable. The dissociated $\alpha\beta$ -dimers are cleared rapidly by renal filtration, which can result in kidney damage.

Looker *et al.*² have now described a genetically engineered haemoglobin molecule incorporating many of the properties required for a red-cell substitute². The oxygen affinity of this molecule compares favourably with that of blood. Dissociation into $\alpha\beta$ -dimers does not occur as the two α -chains are linked together, and this eliminates the renal toxicity. The authors concluded that this engineered haemoglobin is a strong candidate for a safe blood substitute that could be used in the emergency resuscitation of trauma victims.

In our opinion, however, a red-cell substitute must fulfil at least one other important requirement. It must not stimulate bacterial growth, as does ex-

tracellular haemoglobin, because this will promote bacterial infection. In healthy people, most of the iron is located intracellularly as ferritin or as haemoglobin. This iron is normally not available to invading microorganisms. The small amount of extracellular iron that appears in the body is attached to transferrins (host iron-binding and transport proteins), thereby creating bacteriostatic conditions for microorganisms. Extracellular haemoglobin is usually not present in the circulation. Haemoglobin released by lysis of erythrocytes is immediately bound by the serum protein haptoglobin and transported to the liver. Free haemoglobin appears when haptoglobin is saturated. The free haemoglobin is oxidized and dissociates into globin and haem, and the haem molecules are bound by hemopexin and subsequently transported to the liver.

This clearance of extracellular haemoglobin from the circulation is an important defence factor against bacterial infections, as many pathogens possess iron-uptake systems that enable them to use the iron from extracellular haemoglobin. *Neisseria* species, *Haemophilus influenzae*, *Vibrio* species, *Bacteroides fragilis* and *E. coli* are examples of bacteria that are capable of using extracellular haemoglobin^{3,4}. Furthermore, haem-binding proteins of more or less the same molecular mass have been detected in several bacterial species^{5,6}, which could indicate that comparable haem-uptake systems are present in a wide range of microorganisms.

Sickle-cell anaemia, bartonellosis and trauma are examples of conditions where large amounts of haemoglobin may be present in the plasma and in which patients are unusually susceptible to infection⁷. Under these circumstances it is difficult to prove that this unusual susceptibility is due to the liberation of haemoglobin, but the idea that this is reasonable is strongly supported by experimental evidence showing that haemoglobin stimulates the growth of many bacteria in many different hosts⁴. For instance, non-fatal infections with *E. coli* in rats have been converted to fatal infections by adding haemoglobin and the lethal effect is eliminated by the simultaneous administration of haptoglobin⁸. If human recombinant haemoglobin behaves in the same way and promotes bacterial infections, then this will cause considerable problems, particularly in cases of emergency, when patients may have open non-sterile wounds. Therefore, elucidation of the structure–function relationship between the haemoglobin receptor of bacteria and haemoglobin metabolism is needed to provide a basis for understanding the mechanisms used by microorganisms for

acquiring iron from haemoglobin and offer a rational approach for modifying the haemoglobin molecule.

B. R. Otto

A. M. J. J. Verweij-van Vught

D. M. MacLaren

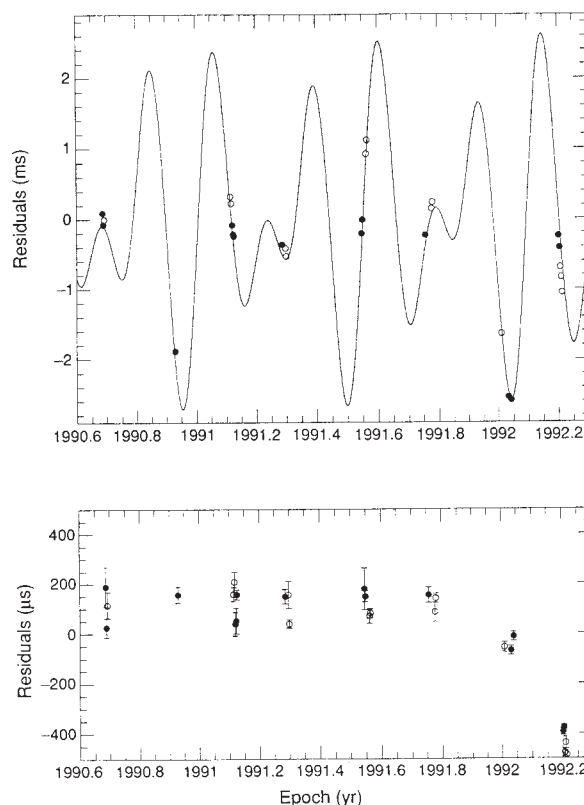
Department of Medical Microbiology,
Vrije Universiteit,
1081 BT Amsterdam,
The Netherlands

1. Ogden, J. E. *Trends Biotech.* **10**, 91–96 (1992).
2. Looker, D. et al. *Nature* **356**, 258–260 (1992).
3. Otto, B. R., Verweij-van Vught, A. M. J. J. & MacLaren, D. M. *Crit. Rev. Microbiol.* **18**, 217–233 (1992).
4. Payne, S. M. *Crit. Rev. Microbiol.* **16**, 81–111 (1988).
5. Coulton, J. W. & Pang, J. C. S. *Curr. Microbiol.* **9**, 93–98 (1983).
6. Lee, B. C. J. *med. Microbiol.* **36**, 121–127 (1992).
7. Kluger, M. J. & Bullen, J. J. in *Iron and Infection: Molecular, Physiological and Clinical Aspects* (eds Bullen, J. J. & Griffiths, E.) 243–282 (Wiley, Chichester, 1987).
8. Eaton, J. W., Brandt, P., Mahoney, J. R. & Lee, J. T. *Science* **215**, 691–692 (1982).

Pulsar's double period confirmed

SIR — We present an analysis of independent data that confirm the perturbation of pulse arrival times of PSR1257+12 presented by Wolszczan and Frail¹ which suggests the presence of a pair of planetary mass companions in orbit around this pulsar. While Wolszczan and Frail's results are particularly convincing in themselves, lingering doubts remain because of an earlier mistaken identification of a planetary-mass pulsar companion^{2,3}.

In September 1990 we added PSR1257+12 to our pulsar timing-array experiment⁴ in the hope that it would provide a stable pulsar clock in a region of the sky widely separated from other members of the array. Observations were recorded at the NRAO 140-foot telescope in Green Bank, West Virginia, every 2 months in 20-MHz bands centred on 800 MHz and 1,330 MHz for 27 days. The data format is a synchronous average over 10 min of the pulse-phase-resolved spectrum with 256 frequency bins and 121 phase bins. A timing model that allowed us to average synchronously was provided by Wolszczan. Reduction of these data to pulse arrival times followed standard procedures developed by us for this experiment⁴. The strong interstellar scintillation of PSR1257+12 provides the necessary gain that allows us to detect this faint pulsar in our



a, Residuals between barycentric arrival times from NRAO 140-ft observations and spin model of ref. 1. Data were recorded at two radio frequencies: 1,330 MHz (solid symbols) and 800 MHz (open symbols). Solid line, residuals expected for the Keplerian perturbations of the two planetary masses discussed in ref. 1. **b**, Residuals between the data and the model presented in **a**. The location of zero on the ordinate scale is arbitrary.

10-min averages with a duty cycle of about 0.20, or for 1 hour out of the 5 that we observe on a given day.

We reduce the arrival-time data to the Solar System barycentre using software developed by the ephemeris group at the Center for Astrophysics⁵. The barycentric reduction and further comparison with the spin model presented by Wolszczan and Frail are conducted in our version of the program TEMPO (see ref. 6). Although there is considerable overlap between our software and that used by Wolszczan and Frail, our results are tightly calibrated by simultaneous pulsar timing-array data on well-studied pulsars such as PSR1855+09, a 5.4-ms pulsar in a 12-day binary system, and PSR1937+21, a 1.6-ms single pulsar, consistent at the microsecond level with published models from data recorded at Arecibo Observatory.

The figure (**a**) displays the day-averaged residuals between our barycentric arrival times and the spin model presented by Wolszczan and Frail. The results for the two observing bands are distinguished by separate symbols. Alignment between the two bands has been effected by a dispersion measure fit after removal of the orbital model. The small adjustment in dispersion measure that we make is not significant owing to our somewhat arbitrary choice of template alignment between the two frequencies. We use the right ascension and declination from the Wolszczan and Frail timing fit. The solid line in the figure is our calculation from

first principles of the residuals expected from the Keplerian perturbations of the two planetary companions whose elements are presented by Wolszczan and Frail¹. The amplitude of these residuals is less than, but close to, one half of the spin period. Our data are insufficiently sampled to achieve a totally independent fit over this interval, but the data show no evidence for pulse number ambiguity between separate days and epochs. The agreement is excellent. Many phases of the 66.6- and 98.2-day orbits are sampled. Use of the VLA position and spin model produces residuals dominated by a roughly one-year sinusoid.

The figure (**b**) also presents day-averaged residuals between the data and the dual-planet perturbation model shown in **a**. Error bars are derived from the individual 10-min residuals on the basis of equal and independent errors; this is not entirely correct owing to the scintillating amplitude of the signal. The agreement then is reasonable over the time range corresponding to the model fit in ref. 1. The residuals for January and March 1992, which are beyond the

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

1. Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
2. Bailes, M., Lyne, A. G. & Shemar, S. L. *Nature* **352**, 311–313 (1991).
3. Lyne, A. G. & Bailes, M. *Nature* **355**, 213 (1992).
4. Foster, R. S. & Backer, D. C. *Astrophys. J.* **361**, 300–308 (1990).
5. Chandler, J. F. *CfA Memo.* 84–81 (1984).
6. Taylor, J. H. & Weisberg, J. M. *Astrophys. J.* **345**, 434–450 (1989).
7. Rasio, F. A., Nicholson, P. D., Shapiro, S. L. & Teukolsky, S. A. *Nature* **355**, 325–326 (1992).
8. Backer, D. C., Fomalont, E. B., Goss, W. M., Taylor, J. H. & Weisberg, J. M. *Astron. J.* **90**, 2275–2280.

observing period of Wolszczan and Frail, are offset by a significant amount, about 170 and 500 μ s, respectively. However, these offsets are not surprising given the number of parameters and uncertainties in Wolszczan and Frail's model. Our modelling of the spin and astrometric parameters shows that most of the growing residual in 1992 can be absorbed by small changes in the spin parameters and position. These uncertainties will need to be reduced before the small orbital perturbations predicted by Rasio *et al.*⁷, which will confirm the planetary hypothesis, can be detected.

A further test was done at the suggestion of a referee by analysing the data with PSRTIME, an independent software package developed by A. Lyne at Jodrell Bank. This package also uses an

independently generated ephemeris, JPL DE118. This comparison required rotating the timing position of Wolszczan and Frail by the known 0.4" shift between the frames of the ephemerides (see ref. 8). A test using PSR1937+21 data showed microsecond agreement between the two analysis packages. Analysis of the Green Bank PSR1257+12 data with PSRTIME using Wolszczan and Frail's timing model also produces good agreement.

D. Backer

S. Sallmen

*Astronomy Department,
University of California,
Berkeley, California 94720, USA*

R. Foster

*Naval Research Laboratory,
Code 4210,
Washington, DC 20375, USA*

Deep-sea amphipod swarms

SIR — Sustained and rapid swimming activity of deep-sea invertebrates is an unusual phenomenon. Even at hydrothermal vents, where food resources are enhanced through chemoautolithotrophic production, the dominant organisms are typically sessile or sedentary, though the swarming shrimp that crawl over surfaces of high-temperature (350 °C) black-smoker chimneys are an exception¹. During *Alvin* dives at the Venture hydrothermal fields² along the East Pacific Rise (9°–10° N, 104° 14–17' W; 2,520 m depth), we encountered a unique swarming behaviour in a previously undescribed pardaliscid amphipod (Fig. 1).

The amphipod swarms are monospecific and form schooling shoals just above

the sea floor at densities sometimes exceeding 1,000 individuals per litre. This concentration of pelagic crustaceans surpasses all reported values from the deep sea that we could find, exceeding Smith's³ estimate of copepod densities at another vent site by three orders of magnitude. The swarms, typically occupying a volume of less than 1 m³, are invariably located immediately downstream of low-temperature (2–8 °C; ambient is 1.8 °C) flows of hydrothermal fluids emanating from cracks in the basaltic sea floor. Orientation of amphipods within a swarm was primarily, but not exclusively, polarized, with individuals facing into the low-temperature diffuse flow.

Using pulsed-dye injections and video

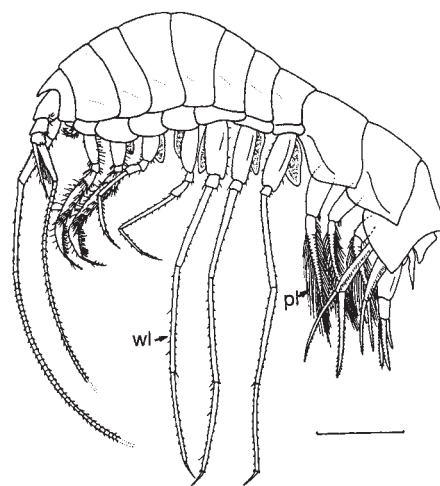


FIG. 2 The swarming amphipods are well adapted for swimming, with strong pleopods (pl) and delicate walking legs (wl), especially pereopods 5–7. Scale bar, 1 mm. Until these observations, pardaliscid amphipods comprised 1% of the known amphipod fauna of vent sites, with 98% of the amphipods belonging to a single lysianassid species⁵.

records, we measured sustained swimming speeds of 5–10 cm s⁻¹ (10–20 body lengths s⁻¹). The amphipods are constrained by a treadmill-like behaviour of net horizontal swimming upstream and passive movement downstream to a cycle length of the order of 2 m or less in an otherwise advective fluid environment. Using a conservative swimming speed of 5 cm s⁻¹, the maximum linear distance covered per day could be greater than 2 km. The swarms appear to be long-lived components of the vent community in that the swarms were first observed during dive operations in May 1991 and were found again at exactly the same sites ($\pm$ 0.5 m) in December 1991.

Not surprisingly, the morphology of the swarming amphipod is adapted for swimming, with well-developed pleopods and delicate walking legs (Fig. 2). Amphipod swarms were observed at numerous sites in association with mussels, clams and tubeworms, as well as in diffuse hydrothermal flows devoid of these megafaunal organisms.

Swarming in crustaceans is mainly reported in shallow-water species where light seems an important cue for swarm maintenance⁴. The significance of swarming in shallow-water taxa is most often associated with reproductive behaviour, feeding and avoidance of visually orienting predators. The adaptive significance of swarming behaviour in vent amphipods which live in complete darkness is not clear. Suspended microorganisms flushed in vent water from subsurface growth chambers may comprise their diet; it follows that individuals must maintain a position where the concentration of food particles is



FIG. 1 High densities of an undescribed pardaliscid amphipod swarm above a mussel (*Bathymodiolus thermophilus*) clump at a vent site. Swarms were found at numerous vent sites near 9° 50' N and at 9° 17' N. Unpublished records of 'crustacean swarms', almost certainly the same amphipod, are referred to in transcripts from *Alvin* dives at vent sites between 10° and 12° N on the East Pacific Rise (G. Thompson and W. Bryan, co-chief scientists; 1988).

sufficient to meet their metabolic demand. Although it is not possible to determine whether the swarming behaviour is more than an expression of optimal spacing of grazing individuals in an advective three-dimensional space, it is clear that the swarming amphipods must play a significant part in the benthic-pelagic coupling of chemosynthetic production.

Cindy Lee Van Dover

Stein Kaartvedt

Stephen M. Bollens

Peter H. Wiebe

Biology Department,
Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543, USA

Joel W. Martin

Natural History Museum of Los Angeles
County,

900 Exposition Boulevard,
Los Angeles, California 90007, USA

Scott C. France

Scripps Institution of Oceanography,
La Jolla, California 92093, USA

1. Rona, P. A. *et al.* *Nature* **321**, 33–37 (1986).
2. Haymon, R. *et al.* *Earth planet. Sci. Lett.* **104**, 515–534 (1991).
3. Smith, K. L. Jr *Limnol. Oceanogr.* **30**, 102–110 (1985).
4. Ambler, J. J. *Plank. Res.* **13**, 1257–1272 (1991).
5. Barnard, J. L. & Ingram, C. *Smithsonian Contr. Zool.* **499**, 1–80 (1990).

Tweaking the magnetosphere

SIR — Kurth's News and Views article¹ captures much of the excitement of the phenomenon of solar coronal mass ejections (CMEs) and their role in triggering terrestrial magnetic storms. But we wish to correct the impression that we do not yet know why some CMEs cause storms and others do not.

The necessary feature for the creation of major magnetic storms is the presence of an intense, long-duration (hours), southward, interplanetary magnetic field^{2–4} somewhere within the high-speed (CME-related) stream structure. These southward fields efficiently interconnect with the Earth's magnetic field and allow energy transfer to the magnetosphere^{5,6}. In a study of 55 (CME-related) shocked high-speed solar wind-stream events occurring in 1978–79, 9 had such intense

long-duration southward fields, 8 had predominantly northward fields and 38 had fields that were primarily in the ecliptic plane or had sufficiently out-of-ecliptic components, but were highly fluctuating in time⁷. Thus, only 9 of 55 shock-led, high-speed (CME-related) streams caused major magnetic storms.

For magnetic storms of lesser intensity, the requirements for the intensity of the southward field component and/or its duration are less⁸. These events are not necessarily associated with high-speed streams or shocks. At the lowest increment of magnetospheric energy, a sub-storm (thought to be an incremental part of a storm) is caused by even weaker and/or shorter-duration southward fields. Substorms are almost never associated

with shocks and/or CMEs but with Alfvén waves⁹ and discontinuities.

The level of the intensity of geomagnetic activity is ordered by the intensity and duration of the southward magnetic field. Velocity plays an important but not dominant role. Thus, CMEs are not a necessary and sufficient condition to cause intense magnetic storms, but long-duration, intense, southward interplanetary magnetic fields are.

Bruce T. Tsurutani

Jet Propulsion Laboratory,
Pasadena, California 91109,
USA

Walter D. Gonzalez

Instituto Nacional de Pesquisas
Espaciais,
Sao Jose dos Campos,
Sao Paulo, Brazil

Emerging cytokine family

SIR — Our colleagues recently reported the cloning of a murine ligand (CD40L) to the CD40 antigen¹. CD40 is a member of a family of at least nine transmembrane proteins², including the low-affinity nerve growth factor (NGF) receptor and two distinct tumour necrosis factor (TNF) receptors that each bind both TNF- α and TNF- β . Although Armitage *et al.* reported that screening of the CD40L sequence against the nucleotide sequence databases revealed no significant similarities¹, we have now found that CD40L is similar to TNF- α and TNF- β (see ref. 3 for a review), suggesting an emerging ligand family parallel to the TNF/NGF receptor family.

CD40L and pro-TNF- α are both type-II membrane proteins and sequence similarity is limited to the carboxy-terminal (receptor-binding) portion of the extracellular domains; in mature TNF- α (and in TNF- β , a secreted protein) the homologous region forms a β -sandwich which trimerizes^{4,5}. CD40L thus presumably shares a similar tertiary structure and may be oligomeric, consistent with ligand-induced receptor cross-linking as a common activation mechanism².

The only other known ligands to mem-

bers of the TNF/NGF receptor family are the neurotrophins (such as NGF), a set of homologous cytokines which bind low-affinity NGF receptor (and several *trk* proto-oncogenes)⁶. Although NGF is a homodimer whose protomer, like TNF- α and TNF- β , has an all- β structure⁷, its topology differs from the TNF topology and no sequence similarity is apparent. The known ligands of the TGF/NGF receptor family thus fall into two structural classes. Because the remaining members of this receptor family (4-1BB, OX40, CD30, CD27 and FaS) reside chiefly on cells of the haematopoietic system, it seems their putative ligands are more likely to resemble TNF than the neurotrophins.

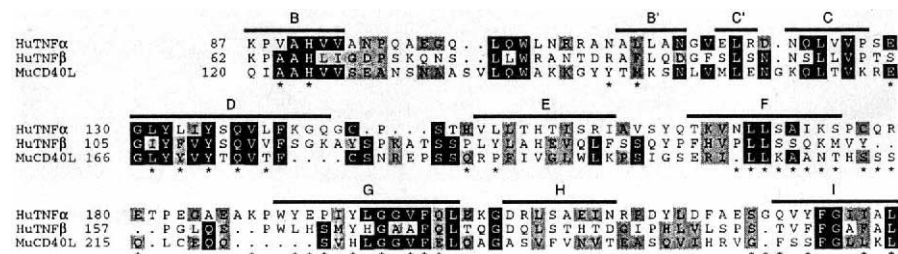
Terry Farrah

Craig A. Smith

Immunex Research and Development
Corporation,
51 University St,
Seattle, Washington 98101, USA

1. Armitage, R. J. *et al.* *Nature* **357**, 80–82 (1992).
2. Smith, C. A. *et al.* *Biochem. biophys. Res. Commun.* **176**, 335–342 (1991).
3. Beutler, B. (ed.) *Tumor Necrosis Factors* (Raven, New York, 1992).
4. Jones, E. Y. *et al.* *Prog. clin. biol. Res.* **349**, 321–327 (1990).
5. Eck, M. J. *et al.* *J. biol. Chem.* **267**, 2119–2122 (1992).
6. Berkemeier, L. R. *et al.* *Neuron* **7**, 857–866 (1991).
7. McDonald, N. Q. *et al.* *Nature* **354**, 411–413 (1991).

1. Kurth, W. S. *Nature* **356**, 18–19 (1992).
2. Burton, R. K., McPherron, R. L. & Russell, C. T. J. *geophys. Res.* **80**, 4204–4214 (1975).
3. Gonzalez, W. D. & Tsurutani, B. T. *Planet. Space Sci.* **35**, 1101–1109 (1987).
4. Tsurutani, B. T., Gonzalez, W. D., Tang, F., Akasofu, S.-I. & Smith, E. J. *J. geophys. Res.* **93**, 8519, (1988).
5. Dungey, J. W. *Phys. Rev. Lett.* **47**, (1961).
6. Gonzalez, W. D. *et al.* *J. geophys. Res.* **94**, 8835–8851 (1989).
7. Tsurutani, B. T., Goldstein, B. E., Gonzalez, W. D. & Tang, F. *Planet. Space Sci.* **36**, 205–206 (1988).
8. Gonzalez, W. D., Gonzalez, A. L. C., Mendes, O. Jr & Tsurutani, B. T. *EOS* **73**, 180–181 (1992).
9. Tsurutani, B. T. & Gonzalez, W. D. *Planet. Space Sci.* **35**, 405–412 (1987).



Region of similarity among human TNF- α and TNF- β and murine CD40L sequences. Identities and similarities between CD40L and each of the TNF sequences are highlighted. Asterisks, residues involved in the trimeric interface of TNF- α ; B–I, β -strands of TNF- α (ref. 4).

That lonesome grail

Sydney Brenner

The Code of Codes: Scientific and Social Issues in the Human Genome Project. Edited by Daniel J. Kevles and Leroy Hood. *Harvard University Press*: 1992. Pp. 397. \$29.95, £23.95.

THE aims of the human genome project are first to obtain a detailed genetic map of the human genome and then to determine the complete sequence of its 3 billion base pairs, Walter Gilbert's "Holy Grail" of biology. The project has always been seen as a large task, requiring a concerted, international effort. The fact that there are now many national genome programmes and several organizations involved in coordinating their activities testifies to the persistence and energy of those who began the project and carried it forward in its early years.

Old controversies

As many readers of *Nature* know, the project has been surrounded by controversy from its very beginning; one of the pioneers, James Watson, has recently departed from the project in uneasy circumstances. Like all crusades, it has attracted passionate supporters, ferocious opponents, believers in the true faith, heretics and even a few schismatic sects. Working on the human genome project is very different from working on the human genome; the latter is exciting science, whereas the former is a banal mixture of metaphysics, politics and, possibly, psychiatry. There have been endless discussions about money to finance the project; now there are also concerns about who owns the genetic information and whether it can be patented. Indeed, there is a vast amount of literature on what one might call 'genomigraphy', the study of the historical, ethical, social, legal and commercial aspects of the project. A cynic once remarked that if all the molecular biologists attending meetings on the human genome project had instead spent their time running a few gels, we might by now have already sequenced a good part of the genome.

Of *The Code of Codes*, it can be said at once that it gives a very balanced cross-section of views on both the scientific aspects of the project and many of the social issues surrounding it. No radical opponents are represented, such as those who argue that there is no need for a special project and that everyday science will complete the work in its own good time, or, at least, that part of the work that is really needed. All the contributors are agreed that the project is here to stay, but several are worried that

it may open a Pandora's box of undesirable consequences. In the introduction, Daniel Kevles points to the strong connection between the genome project and eugenics. Classical eugenics was a social programme for improving humans through their heredity; its proponents believed that social degeneration was produced by bad genes and that if these could be eradicated by selective breeding, society would be cured of its 'genetic diseases'.

This form of eugenics has long been discredited and has now been replaced by a medical programme to help individuals and their families to deal with their private genetic diseases. There is a superficial resemblance here to what happened for infectious diseases. These were once dealt with by public health programmes of sanitation and vaccination, but, today, in advanced societies, infections are very much a matter for individual medical therapy by antibiotics. The comparison ends here, though, because while everybody knows that infectious diseases are caused by bacteria and viruses, nobody believes that social diseases have genetic causes.

But many of the contributors consider the social consequences of genome analysis. Evelyn Fox Keller, in particular, is concerned that the 'new eugenics' will be no better than the old, with the same ideology of subsuming culture under biology, except that now the ideology is powerfully supported by the scientific successes of molecular biology. Such worries are realistic. The common genetic diseases, such as cystic fibrosis and muscular dystrophy, are caused by mutations in single genes, and can readily be analysed genetically. The mapping of these mutations makes it possible to screen for the diseases and families can then practise zygote selection. Furthermore, the identification of the responsible genes may lead to gene therapy, and even perhaps to the replacement of their function.

All of this fits well into the structure of modern medicine, where individuals are given information and can use it to make informed decisions. It is safe to say that as long as the new knowledge about these common genetic diseases is delivered through the conduit of current medical practice, it will not create any new ethical problems. More problema-

tic, though, is the question of genetic propensity to disease, where the environment may play a commanding role. Unlike contemporary medicine, which deals with the products of historical pathogenesis, 'predictive medicine' claims to tackle the roots of future disease processes. This medicine is completely embraced by Leroy Hood in his essay entitled "Biology and Medicine in the Twenty-First Century", an optimistic panegyric of what biotechnology is going to accomplish in the next 20 years. He believes that the most important area of DNA diagnostics will be in the "identification of genes which (*sic*) predispose individuals to disease", and he mentions cardiovascular, neurological and autoimmune diseases. Perhaps he subsumes psychiatric illness under neurological disease; other contributors to the book certainly include it. These diseases are all said to be polygenic, that is they involve more than one gene. It is not clear, however, whether each disease is a collection of many different single gene defects or represents a unique syndrome that is the outcome of several predisposing mutations triggered by a particular environment. This is mainly a question of objective diagnosis and classification. Indeed, much of the confusion in genetic studies of mental disease arises from difficulties in diagnosis. In any event, genetic analysis can help us considerably to tease out the different single gene entities from disease collections, and thus provide an objective classification, as has happened in cardiovascular disease — a conglomerate disease if there ever was one.

Polygenic diseases pose an entirely new question, because here the future appearance of disease can be predicted only as the probability of its occurrence in any given individual. If there is no remedy, then this genetic information is of no use to the individual and there is no rational medical basis for establishing it. This 'probabilistic medicine' is of use only to institutions such as government agencies and insurance companies, and the conflict between individual and institutional interests needs to be properly resolved in our society or else, as Dorothy Nelkin points out in her important essay on this subject, we risk "creating a genetic underclass".

New genetics

The underlying science of the project is of fairly recent development. DNA cloning and DNA sequencing were invented in the mid-1970s and radically altered the field of genetics. Geneticists were effectively liberated from the tyranny of the breeding cycles of animals, and, in principle, the main tasks of classical experimental genetics — the definition of genes by mutation and complementa-

tion — could be pursued in any organism by cloning and sequencing. Genetic maps, relying on breeding experiments, are required only when one wishes to operate on the organism in the real world, such as in the study of human genetic disease or in the improvement of plants and animals, but, even here, the new techniques reveal new sources of genetic variation in the form of DNA markers that greatly accelerate map construction.

In his essay entitled "A Vision of the Grail", Walter Gilbert looks at how molecular biology in general, and the human genome project in particular, will change biology. He believes that most of the present style, if not content, of today's experimental biology will wither away, and we will purchase all the tools and information as kits and services. Sequences — even the grail sequence — will be available through computer networks and the "science will have moved on to the problem of what a sequence means, what the gene actually does". He published this view last year in *Nature* (349, 99; 1991), and I do not know why it aroused such a violent response; true, it is a somewhat arrogant way of putting it, but after many years of hard work I, for one, look forward to an elegant life of theoretical contemplation of sequences, punctuated by the odd delightful, well-designed experiment. Of course, the lament of the older molecular biologist is that the kit revolution has already happened; most young gene hackers would not know how to go about constructing a vector or making a mutant, and even those who claim to be

discovering new human genes find them in commercial libraries.

The new genetics in fact involves a return to a much older tradition in science, that of observation and measurement, a kind of paradigm regained rather than one lost. The study of genomes differs as much from experimental biology as astronomy differs from experimental physics; 'genonomy' seems an appropriate name for this new science. For us genomers, who chart the genetic heavens, the invention of sequencing can be compared to Galileo's invention of the telescope, and like astronomers, we can also look far back in time, reconstructing the genes and genomes of long-vanished organisms. We do not have hypotheses, and we have no experiments to perform, so we are unlikely to command the attention of those who dispense research funds on a strict Popperian basis. We do not want to stop experimental biology, because that is needed for the analysis of biological function, which tells us the meaning of our sequences. But those many hedgehogs interested in finding out everything about one thing should tolerate the few foxes who wish to find out a little about everything. In studying the human genome, much will be discovered about the evolution of life and living systems, and if, as the book tries to show, there are fears, there is also hope that this knowledge will also benefit humanity. What more can one want? □

Sydney Brenner is in the MRC Molecular Genetics Unit, Hills Road, Cambridge CB2 2QH, UK.

Making no apologies

Martin Gardner

The Art of Mathematics. By Jerry P. King. Plenum:1992. Pp. 313. \$24.50.

CREATIVE mathematicians seldom write for outsiders, but when they do they usually do it well. Jerry King, a professor at Lehigh University, Pennsylvania, is no exception. His informal, nontechnical book, as its title implies, is organized around what Bertrand Russell called the "supreme beauty" of mathematics — a beauty "capable of a stern perfection such as only the greatest art can show."

King sharpens C. P. Snow's famous distinction between the two cultures by making it a distinction between an M-culture (those who know mathematics) and an N-culture (those who don't). Widespread ignorance of mathematics, he is convinced, springs mainly from incompetent instructors of the young, often people who actively dislike what they teach. King recalls amusing ways in which he was damaged by such teachers until one fine day, almost by accident, he took a course in which the beauty of calculus struck him like the light that felled Saul on his way to Damascus.

King's tour of his topic begins with the counting numbers, then takes the reader painlessly through their generalizations to the rationals, the reals, through imaginary and complex numbers, and into the wonders of elementary geometry. He is good at conveying the supreme importance of the concept of limit, and how it underlies both branches of calculus. A drawing depicts the differential and integral calculus as two sides of a great arch, each a mirror image of the other, held together by a capstone that is the fundamental theorem of calculus. There is a long chapter on the philosophy of aesthetics, and another on the reluctance of a mathematical aristocracy to reform mathematical teaching.

King has nothing to say about recent developments such as computer science, fractals, chaos theory, or topology with its spinoffs of graph and knot theory. To go into these trends, however, would have defeated his main purpose — to introduce low-level mathematics to readers who have learned to hate the subject.

King avoids the trap of thinking that the discipline is infected by the uncertainty of science. Pure mathematics, he makes clear, "lives" inside formal systems. Given a system's assumptions, theorems are not only unassailable, they are the only kind of certainty we mortals have. King's answer to Pilate's question



The South American tree *Cinchona* (Rubiaceae) is widely cultivated for the dried bark of its stems and roots, the source of quinine and other antimalarial alkaloids. There are more than 10,000 species worldwide. Shown here is *Cinchona ledgeriana*, which is grown commercially in the tropics. The picture first appeared in *Medicinal Plants* by R. Bentley and H. Trimen (1880); it is now reproduced in *Plants in Cardiology* by A. Hollman, published by the British Medical Journal at £5.95 (pbk).

is "Truth is what you find at the end of a correct chain of arguments."

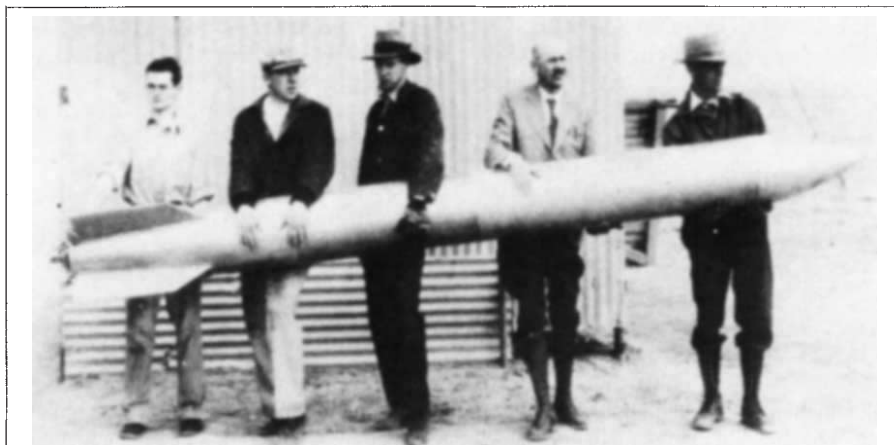
The author quotes Einstein: "The reason why mathematics enjoys special esteem, above all other sciences, is that its laws are absolutely certain and indisputable, while those of all other sciences are to some extent debatable." Two plus two cannot be anything but four, but when one applies this tautology to the world one has to invoke what the philosopher Rudolph Carnap called "correspondence rules". These are assumptions such as that a pebble, a cow or a star is taken as 1. If such equivalence is not assumed, then $2 + 2 = 4$ is violated by, for example, the fact that two drops of water added to two drops makes one big drop. Nothing could be simpler than this obvious distinction between pure and applied, yet some philosophers and mathematicians, frightened by the word 'truth', seem unable to grasp it.

King shares physicist Eugene Wigner's puzzlement over the "unreasonable effectiveness" of mathematics. This puzzlement puzzles me. Not only is the Universe mathematically structured, it is made entirely of mathematics. Matter consists of fields and their particles, which are not made of anything except equations. What is so mysterious about the application of these equations to a Universe from which our minds, in turn made of mathematics, originally extracted them? If mathematical patterns are not discovered, but are entirely human creations like poetry and music, then their powerful effectiveness in science and technology is indeed what Wigner called an undeserved miracle.

King is particularly fond of the poetry of Robert Frost. His poems "stay with you because of their simplicity: you come back to them for their depth . . . Frost brought poetry to the multitudes; let's now bring them mathematics." The Frosts of mathematics have yet to appear, but King is "certain as sunrise" that they will some day revolutionize mathematical teaching.

The author has a romantic vision of himself as "an aging gunfighter, walking alone the mean streets of academe". His fire is aimed at a conservative elite too self-absorbed in their specialized research to be much concerned about the public's growing mathematical illiteracy. The dull "cinder-block textbooks" get "clunkier". Although their diagrams are splattered with colour, somehow the splendour of ideas behind the diagrams remains as invisible as the textbook authors. Meanwhile the M- and N-cultures "fly apart like galaxies on opposite sides of an expanding universe". □

Martin Gardner is at 110 Glenbrook Drive, Hendersonville, North Carolina 28739, USA.



Holding tight — the pioneering rocket experimentalist Robert H. Goddard (second from right) displaying a liquid-propulsion rocket used in a flight in 1932. Picture from *History of Liquid Rocket Engine Development in the United States 1955–1980* edited by S. E. Doyle. American Astronautical Society, \$50 (hbk), \$35 (pbk).

Differentiating diseases

Denis Mollison

Infectious Diseases of Humans: Dynamics and Control. By Roy M. Anderson and Robert M. May. Oxford University Press: 1991. Pp. 757. £50, \$95 (hbk); £22.50, \$45 (pbk).

THROUGHOUT history, many more people have been killed by infectious diseases than by wars. One spectacular example is the conquest of the Americas following Columbus's arrival 500 years ago. The sudden collapse of the great Aztec and Inca empires could not have happened without the help of smallpox and other 'European' diseases. Often spreading faster than the colonists themselves, these microscopic invaders in a few years reduced the native population to less than a tenth of what it had been previously.

Despite such evidence of the enormous impact of infectious diseases, it is little more than 100 years since the germ theory of disease won general acceptance through the work of Pasteur, Koch and others. Mathematical modelling of disease began in the early years of this century with, for instance, Hamer and Soper's work on the basic dynamics of outbreaks and recurrent cycles of endemic diseases such as measles. Over the past 30 years or so, the literature of theoretical epidemiology has gathered pace, rather as though it were itself a new and virulent disease.

As Anderson and May point out, much of this work has become too detached from data, as well as from public health policy. Their primary aim in this book is "to show how simple mathematical models of the transmission of infectious agents within human communities

can help to interpret observed epidemiological trends, to guide the collection of data towards further understanding, and to design programmes for the control of infection and disease".

Anderson and May divide infectious diseases into two broad classes, requiring different kinds of models, according to whether the diseases are caused by microparasites or macroparasites, the key distinction being that the former reproduce directly in the human host. For each class, they begin with simple static models for a stable situation of endemic disease and with simple dynamic models, before introducing important complications such as age dependence, genetic, social and spatial heterogeneity, and the problems of control. In each case, the modelling is clearly motivated by real diseases, and the discussion tied closely to clear presentations of data across the range of human infectious diseases.

This structure works beautifully, avoiding repetition and bringing out the similarities between diseases whose behaviours are superficially very different. There is a welcome emphasis on starting with simple models, so as to avoid, in the authors' words, becoming lost in "a snowstorm of parameters". It is still insufficiently recognized in mathematical biology that the 'flexibility' of a model with many parameters often amounts to meaninglessness, in that such a model can be fitted to anything; as an eighteenth-century mathematician once put it: "give me five parameters and I will draw you an elephant, six and I will have him wave his trunk".

The authors' mathematical approach is based on differential equations. Although possibly off-putting in appearance, at least to the biological and medical scientists at whom the book is aimed, these equations are straightforward and well-understood mathematical tools, and allow the authors to consider

a marvellously thorough range of aspects of disease, from comparison of different vaccination policies to the coevolution of host and parasite populations.

Differential equations in biology do have their problems, however. Their air of apparent precision (mimicking the genuine precision of, say, the law of gravity, $\ddot{x} = -gx$) can mislead the unwary. For instance, it is a common and reasonable simplification to represent the rate at which infected individuals recover as being proportional to the number of infected individuals: in its simplest form, $\dot{Y} = -\mu Y$. Unfortunately, this form carries the implicit assumption that the duration of the infectious period has a specific distribution, the exponential, which is a poor approximation in most cases. Although this unwanted precise assumption generally has little effect on the dynamics of the model, it can make a difference to the stability or otherwise of the endemic state.

Sorting out robust conclusions from those that are sensitive to unreliable modelling detail is a delicate task for the mathematician. Regrettably, biologists often use this overprecision as an excuse to dismiss the mathematical approach to disease dynamics (or, equally undesirably, they swallow specific differential equations uncritically). May and Anderson are, of course, well aware of such problems, and are full of insights into the general qualitative truths that lie behind specific details. A nice example is their summary of the timescales of dynamics in terms of those of the disease itself (its 'generation gap', D) and the human population (its mean lifetime, L). They describe the initial epidemic outbreak, the regrowth of susceptibles that follows, and the later coupling of infected and susceptible individuals in endemic oscillations, as being on timescales of D , L and their geometric mean $\sqrt{DL}$, respectively.

Even Anderson and May can get caught up in the obfuscations of the differential-equation approach, as in their basic consideration of equilibrium in chapter 4. Here they establish a number of simple and appealing results, relating basic epidemic and population parameters, as approximations in specific cases, when slight modifications can be shown to hold more generally by simply considering the logically possible flows between disease states.

At a deeper level, many such simple relationships become poor approximations when one gets into the complexities of population heterogeneity. Some of these complexities may require models framed in terms of individuals; a basic imprecision of differential equations is their treatment of population as a continuous mass, which makes it difficult to use them to model spatial or social

contact structures. Corresponding data studies at the individual and local level will also be required. Recent improvements in computer technology have helped here by providing better databases, and by opening up new possibilities for the use of computer-intensive methods in the analysis of data.

This is a masterly survey of a wide range of aspects of the dynamics of human diseases. It is difficult to see how such width could have been achieved other than by using simple differential-

equation models. But these models rarely provide the last word; when one considers the details of populations, it is often unclear what level of accuracy has been achieved. The book should therefore be seen as setting a rich agenda for further research rather than as a final solution. □

Denis Mollison is in the Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK.

Taming the oceans of the world

Donald S. McLusky

Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Oceans. By K. H. Mann and J. R. N. Lazier. *Blackwell Scientific*: 1991. Pp. 466. £24.95, \$46.95 (pbk).

WHEN A. G. Tansley introduced the term "ecosystem" in 1935, he stressed the interaction between living and non-living components. Over the years, biologists have come to use the word mainly for studies of the energy transfers within food webs. These studies have been useful for coordinating research and as models for explaining ecological principles. At the same time, physical techniques have become more sophisticated and scientists can now study processes in the world's seas using satellite imagery and other ways of collecting data quickly and directly.

Mann and Lazier claim that the time has come to rejoin these two branches of study, with marine ecology emerging as an integrated discipline concerned with the physics, chemistry and biology of the sea. The relevance of this new marine ecology has probably never been greater: the physical processes underlying large-scale biological phenomena such as production at upwellings of fronts are now better understood; measurements of biological and physical parameters can be made at truly comparable levels; the need to understand the influence of marine ecological processes on world climate is urgent; and a study of fundamental ecological processes is important in ocean management.

The book is imaginatively divided into three sections on the basis of scale. The first is concerned with processes on a scale of less than 1 km. The authors admit that grappling with the physics of turbulent flow is hard labour for many biologists but claim that it is essential for understanding contemporary marine ecology. From this beginning, the reader approaches the vertical structure of the ocean. Here the authors acknowledge

that the long-term understanding based on Sverdrup's observations has been confirmed by modern technology. From a combined physical and biological understanding, four general patterns of primary production in the oceans can now be recognized.

In the second section, the authors discuss processes on a scale of 1–1,000 km, and include detailed accounts of coastal upwelling based on studies of the Canary, Peru, California and Benguela currents, and studies of fronts in coastal waters. The authors freely confess that the latter field is new and that there is so far no consensus among researchers.

The third section covers processes on a scale of thousands of kilometres, including the circulation of the ocean basins, and the biological consequences of variability in ocean circulation, including events from the North Sea and the Pacific. In the final chapter, the authors look at the oceans and global climate change. They point out the crucial role of the oceans in carbon dioxide balance — the oceans and their sediments contain almost all the carbon dioxide ever issued into the atmosphere. They also consider the many complications in assessing the role of phytoplankton in pumping carbon into the interior of the oceans and, after summarizing the many alternatives, wisely conclude that it is, as yet, impossible to be certain what the overall effects will be on the oceans or the atmosphere.

The book is well written and the authors' approach works well. They give a clear plan in the first chapter of their aim to unite varied disciplines, and each chapter is complemented by a simple summary, as well as 'boxes' to provide supplementary reading or formulae. I am sure that the book will be adopted as a standard text for many marine science courses. More importantly, it highlights enough unanswered questions to motivate a new generation of researchers. □

Donald S. McLusky is in the Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, UK.

Quantification of the disorder in network-modified silicate glasses

I. Farnan*, P. J. Grandinetti†, J. H. Baltisberger†, J. F. Stebbins*, U. Werner†, M. A. Eastman† & A. Pines†

* Department of Geology, Stanford University, Stanford, California 94305, USA

† Materials Science Division, Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA and Department of Chemistry, University of California, Berkeley, California 94720, USA

Local order in silicate glasses has been observed by many experimental techniques to be similar to that in crystalline materials. Details of the intermediate-range order are more elusive, but essential for understanding the lack of long-range symmetry in glasses and the effect of composition on glass structure. Two-dimensional ^{17}O dynamic-angle-spinning nuclear magnetic resonance experiments reveal intermediate-range order in the distribution of inter-tetrahedral (Si–O–Si) bond angles and a high degree of order in the disposition of oxygen atoms around the network-modifying cations.

THE difference between a glass and a crystal lies in disorder present in the intermediate-range glass structure that eliminates long-range translational symmetry. Characterization of this disorder is an important experimental objective because it is a critical test of the accuracy of models of glass structure. In network-modified silicate glasses, the continuous disordered network of SiO_4 tetrahedra is presumed to be disrupted by modifying cations which create nonbridging oxygens (oxygen bonded to one silicon atom). Two principal sources of disorder are thought to be this disruption of the network and the range of bond angles (Si–O–Si) between network-forming cations. It is well established experimentally that silicon and oxygen are ordered locally in silicate glasses, and that SiO_4 tetrahedra are the basic structural units just as they are in low-pressure crystalline silicates. From extended X-ray absorption fine structure (EXAFS) studies of modifier cations, we know that they too are regularly coordinated by oxygen^{1,2}, having coordination polyhedra and bond lengths similar to those in crystalline silicates. Isotopically substituted neutron scattering has also shown that ordering associated with modifier cations extends beyond the first coordination sphere, by detecting strong correlations between Ca–Ca as well as Ca–O distances in CaSiO_3 glass^{3,4}. This is consistent with ^{29}Si nuclear magnetic resonance (NMR) studies of silicate glasses, which show that the distribution of nonbridging oxygens is not random, being close to binary⁵; the deviation from a binary distribution depends on the electronegativity of the network modifier and on the glass transition temperature⁶. Taken together, these experimental data indicate considerable order associated with network modification. Quantification of the remaining disorder associated with variations in bridging angles between network-forming cations is therefore important. For unmodified SiO_2 glass, the X-ray scattering radial distribution function⁷ and correlations between ^{29}Si NMR chemical shifts and average Si–O–Si bond angles^{8,9} have been used to derive the inter-tetrahedral (Si–O–Si) bond-angle distribution. Both of these methods encounter problems when extended to compositionally more-complex modified glasses. Almost all technologically or geochemically important glass compositions are more complex than SiO_2 . The ability to describe and quantify the microscopic structure of these materials (beyond the first coordination sphere) as a function of composition will considerably advance the understanding of their structure and properties.

Volumetrically, silicate glasses are dominated by oxygen anions, yet despite this the structure of silicate glasses has been studied almost entirely from the perspective of the cations and their coordination. For example, X-ray scattering experiments

are most sensitive to scattering from cations (network-forming and network-modifying) that are heavier than oxygen anions, EXAFS experiments concentrate on network modifiers such as Na^+ or Ca^{2+} , and ^{29}Si NMR experiments specifically observe the signal from the network-forming cation. Here we investigate the local environments of these oxygen anions. As oxygen is the connecting atom between locally ordered SiO_4 tetrahedra, and between silicon and network-modifying cations which are also thought to have locally ordered environments, the intermediate-range disorder in the glass will be reflected in the range of environments exhibited by these oxygen atoms. Oxygen-17 NMR is a sensitive and direct way to characterize these interconnections. Spectral resolution in previous ^{17}O NMR studies of silicate glasses^{10,11} was seriously reduced by anisotropic interactions between ^{17}O nuclei and electric field gradients (EFGs) at oxygen sites. Line broadenings caused by a distribution of oxygen environments contained information on the disorder in the glass but were convolved with the anisotropic quadrupolar broadening, which could not be removed by the magic-angle-spinning technique¹². In contrast, two-dimensional dynamic angle spinning (DAS)¹³ can produce high-resolution ^{17}O NMR spectra of crystalline silicates. Similar experiments on silicate glasses should provide isotropic ^{17}O chemical and quadrupolar shifts that have the potential to identify bridging and nonbridging oxygens, discriminate between oxygens bonded to different network modifiers and provide details of the ordering of those modifiers around the nonbridging oxygens. Additional insight into the structure of these glasses is available in the anisotropic second dimension of the two-dimensional (2-D) NMR spectrum, where anisotropic line shapes can be correlated with oxygen site symmetries.

We present ^{17}O NMR results for potassium and potassium–magnesium silicate glasses that indicate a high degree of ordering of network-modifying cations around nonbridging oxygens even when two network modifiers are present. Although ordering of single network-modifying cations in silicate glasses may be expected from previous neutron-scattering experiments^{3,4}, substantially more order is implied here as two different cations order about the nonbridging oxygens. In both glasses there is a high degree of uniformity in nonbridging oxygen environments, with oxygens that bridge SiO_4 tetrahedra having a greater range of environments than those that bond to network modifiers. Presumably, without the constraint of long-range symmetry the structure can relax locally around the modifier, with most of the disorder being assumed by the distribution in intertetrahedral Si–O–Si angles. This bond-angle distribution can be determined by combining the information on the oxygen

site symmetry available from the 2-D NMR experiment with correlations between the asymmetry and magnitude of the EFG at the bridging oxygen and the Si-O-Si bond angle. Such detailed structural information should provide accurate data for modelling and allow experimental evaluation of the effect of pressure and temperature (glass transition temperature) on vitreous SiO₂ and modified silicate glasses of complex composition. In addition, this method probes the network-forming anion and so should be applicable to any oxide glasses.

Experimental methods

NMR frequency shifts introduced by magnetic or electric interactions due to local atomic structure are generally anisotropic: the shift depends on the orientation of the atomic site with respect to the external magnetic field. In powdered or amorphous solids, all possible orientations of sites are present, and the superposition of resonances from all orientations produces broad anisotropic line shapes. To achieve high-resolution NMR in this situation, the broadening can be removed by rapid sample spinning at specific angles to the magnetic field. The relative orientation of the site and the magnetic field become time-dependent, and the broadening interaction is averaged away either completely or to a single scalar (isotropic) value. For broadening interactions small enough (compared with the interaction with the external magnetic field) to be described to first order, such as chemical shift anisotropy, spinning the sample about a single axis inclined at an angle of 54.7° to the magnetic field (magic angle spinning) will result in a high-resolution spectrum. In general, however, the interaction between the quadrupole moment of ¹⁷O and the EFGs at oxygen sites is large enough that second-order broadenings are substantial and they can only be completely removed by spinning the sample about two axes¹², as in the recently developed DAS NMR technique^{13–15}. DAS is intrinsically a two-dimensional experiment and provides a correlation between high-resolution isotropic spectra in the ω_1 dimension, where all anisotropic broadening is removed (each oxygen site has a characteristic isotropic chemical and second-order quadrupolar shift), and anisotropic site symmetry information in the ω_2 line shape.

We prepared two silicate glasses of composition K₂Si₄O₉ and KMg_{0.5}Si₄O₉ with nominal ¹⁷O enrichments of 35 and 41%, respectively. These compositions were chosen to test the ability of the DAS NMR technique to distinguish oxygens bonded to different network-modifying cations as well as between bridging and nonbridging oxygens. Oxygen-17 DAS spectra were acquired for both samples using a home-built probe¹⁶ in commercial wide-bore magnets with fields of 9.4 and 11.7 T. The 2-D ¹⁷O NMR spectrum of K₂Si₄O₉ glass is shown in Fig. 1a. Resonances characteristic of bridging (6 p.p.m.) and nonbridging (64 p.p.m.) oxygens in the glass are clearly resolved in the isotropic spectrum (projection of the 2-D spectrum onto the ω_1 axis) and the 2-D plot. The 2-D plot also shows structure in the bridging oxygen resonance in the anisotropic dimension (ω_2) with the line shape spreading out towards higher frequency in ω_1 . The isotropic linewidths, which are roughly an order of magnitude greater than those obtained for ¹⁷O DAS NMR in crystalline silicates, are due solely to distributions in the environments of the bridging and nonbridging oxygens and so directly reflect the disorder associated with each oxygen type.

Ordering of network modifiers

The range of environments displayed by a nonbridging oxygen will reflect both the distribution of network modifiers throughout the sample and the regularity of the individual network-modifier environments. The isotropic ¹⁷O DAS NMR resonance of the nonbridging oxygen in K₂Si₄O₉ glass (Fig. 1a) at 64 p.p.m. is 17 p.p.m. wide (full width at half maximum, FWHM). A distribution of isotropic quadrupolar and chemical shifts for each oxygen site is responsible for this width. In previous studies of crystals, the chemical and quadrupolar shifts of unique sites

were separated by taking advantage of the inverse magnetic field dependence of the second-order quadrupolar interaction¹⁷ and doing the experiment at two magnetic fields. The chemical shift may then be determined by extrapolation to infinite magnetic field where the quadrupolar shift becomes zero. The exact chemical shift distribution in a glass cannot be derived by extrapolation, or mapping, to infinite magnetic field in this way without prior knowledge of the relationship between the chemical shift and quadrupole parameters for the oxygen sites. The change in FWHM with magnetic field should however allow us to obtain a FWHM for the distribution of chemical shifts. The nonbridging line in K₂Si₄O₉ glass at 11.7 T (not shown) is 15 p.p.m. wide, and this, combined with the 9.4-T data, indicates a range of chemical shifts of 13 p.p.m. This range is narrow compared with that of nonbridging oxygens in crystalline silicates¹⁸ where ¹⁷O chemical shifts of oxygens bonded to single

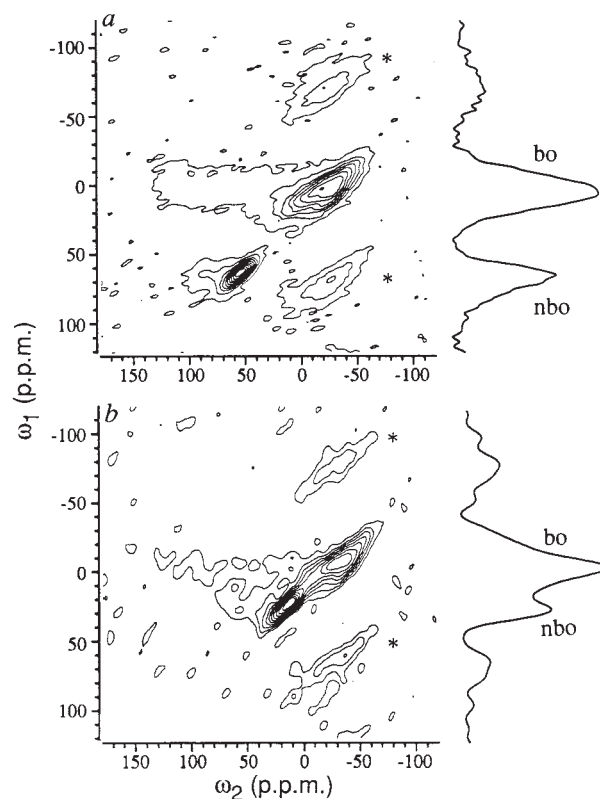


FIG. 1 Two-dimensional ¹⁷O DAS NMR spectra and projections onto the isotropic axis (ω_1) of modified silicate glasses at 9.4 T (54.245 MHz): a, K₂Si₄O₉; b, KMg_{0.5}Si₄O₉. The resonances of bridging oxygens (bo), nonbridging oxygens (nbo) and sidebands (*) that arise from spinning slower than the width of the static anisotropic line are marked. The spectrum is obtained by spinning the sample sequentially about two axes inclined at angles θ_1 (37.38°) and θ_2 (79.19°) to the external magnetic field. The angles are chosen so that the anisotropic evolution of the ¹⁷O spin magnetization at θ_1 is the inverse of that at θ_2 and is thus completely cancelled during evolution at the second angle. By incrementing the evolution time t_1 at angle θ_1 and detecting each NMR signal $S(t_2)$ while spinning about angle θ_2 , we obtain a 2-D interferogram $S(t_1, t_2)$ which can be Fourier-transformed and phase-corrected to produce a 2-D frequency spectrum containing an isotropic dimension (ω_1) and an anisotropic dimension (ω_2). This produces anisotropic line shapes in ω_2 characteristic of spinning about an angle of 79.19° to the magnetic field. These are due almost entirely to the interaction between the quadrupole moment of ¹⁷O and the EFG at the oxygen sites, and have shapes that depend on the symmetry (characterized by the asymmetry parameter, η) and widths that depend on the magnitude of the EFG. Selective pulses that excited the ¹⁷O central transition and delays of 4 s were used to collect ~50 t_1 points. The t_1 dimension was zero-filled to 256 points and a 200-Hz gaussian line broadening was applied in the t_2 dimension before Fourier transformation of a 256 × 256 data matrix.

modifier cations can range up to 27 p.p.m. in wollastonite (CaSiO_3) and 23 p.p.m. in forsterite (Mg_2SiO_4). It does seem that there is a greater range of (discrete) nonbridging oxygen environments in crystalline silicates than in the glasses studied here, although the variation in ^{17}O chemical shift with field strength of the modifier cation may not be the same. The implication is that the lack of long-range symmetry constraints in a glass allows local structural relaxation around network modifiers and produces more-uniform nonbridging oxygen environments.

The disorder associated with nonbridging oxygens might be expected to increase if we were to introduce an additional modifier into the tetrasilicate composition glass. Figure 1b shows the spectrum of $\text{KMg}_{0.5}\text{Si}_4\text{O}_9$ glass at 9.4 T. The NMR line for the nonbridging oxygen shifts by 40 p.p.m. to lower frequency while retaining essentially the same width as the nonbridging oxygen peak in $\text{K}_2\text{Si}_4\text{O}_9$. Only one peak is observed, so there are no separate nonbridging oxygen environments due to K and Mg modifiers, nor any clustering of K and Mg. The similar width of the nonbridging oxygen peak in both glasses implies a high degree of ordering of K and Mg. Each nonbridging oxygen in the $\text{KMg}_{0.5}\text{Si}_4\text{O}_9$ glass must have a similar bonding environment, with the linewidth caused by similar variations in bond length and angle as in the single modifier glass. A random distribution of K and Mg, on the other hand, might be expected to produce a broad peak ranging from 64.0 to 25.0 p.p.m. (comparing the shifts of K and Mg nonbridging oxygens in crystalline materials^{19,20}). In crystalline $\text{K}_2\text{Si}_4\text{O}_9$ (ref. 21) each nonbridging oxygen is in fivefold coordination with one silicon and four potassium neighbours. If the nonbridging oxygens in the glass have similar environments to the crystal an ordered substitution of one Mg^{2+} for two K^+ about each nonbridging oxygen is implied. For a K/Mg ratio of 2, the only way to produce an ordered arrangement of K^+ and Mg^{2+} in the glass is to have an original coordination of four K^+ about each nonbridging oxygen in $\text{K}_2\text{Si}_4\text{O}_9$. This fivefold coordination of the nonbridging oxygens in alkali silicate glasses is also predicted by recent molecular dynamics simulations²². The ordered substitution of Mg^{2+} in the tetrasilicate glass will reduce the nonbridging oxygen coordination number from five to four. We determine a change in the average chemical shift of 26 p.p.m. for the nonbridging oxygen site, from data at 11.7 T. This change is similar to that observed for changes in chemical shift with oxygen coordination number of cations such as Si or Al, but in the opposite direction: lower coordination is more shielded.

The uniform and ordered nonbridging oxygen environments observed here, together with experiments on the network modifiers themselves^{1,3,4} that show 'unexpected' local order

around these cations, indicate that the configuration 'frozen in' as a silicate liquid goes through the glass transition is one that corresponds to a minimum energy and high order for the network modifiers and nonbridging oxygens.

Inter-tetrahedral bond angles

In Fig. 1a and b, the 2-D spectrum of the bridging oxygen shows structure in the anisotropic dimension, with a systematic change in the anisotropic line shape across the isotropic line from low to high frequency. Slices parallel to the ω_2 axis (Fig. 2) show this change in line shape to be due to an increase in the asymmetry parameter (η) with increasing frequency in ω_1 . The parameter η describes the symmetry of the EFG at the oxygen site. A physical interpretation of this behaviour is that low values of η (0 represents cylindrical symmetry) correspond to Si-O-Si bond angles close to 180° , and as the bond angle decreases the deviation from cylindrical symmetry increases, thus increasing η .

In contrast to η , the isotropic quadrupolar shift (to low frequency from the chemical shift) is almost entirely determined by the magnitude of the EFG (its asymmetry has a very small effect). Oxygens shifted to lower frequency in ω_1 have smaller values of η and those shifted to higher frequency in ω_1 have larger values of η . This means that the oxygen EFG is a maximum for Si-O-Si bond angles of 180° , and decreases with bond angle. This interpretation is supported by *ab initio* molecular orbital calculations²³ of the variation of the oxygen EFG and η with the Si-O-Si bond angle (α) in the molecule $\text{H}_3\text{Si-O-SiH}_3$. Analysis of these calculated values shows a simple relationship: $\eta = 1 + \cos \alpha$. These calculated values of η are plotted along with this function in Fig. 3a. Two experimental values of η determined from careful fitting of bridging oxygen ^{17}O MAS NMR spectra of cristobalite (SiO_2) and wadeite $\text{K}_2\text{Si}_4\text{O}_9$ (a high-pressure polymorph of $\text{K}_2\text{Si}_4\text{O}_9$) also fall close to this line.

The observed line shapes in ω_2 may be analysed using a multi-parameter least-squares fitting procedure to extract η , the EFG and chemical shift for each slice (Fig. 2). The result of fitting the slice through the maximum of the bridging oxygen peak gives a value for the most probable angle in the Si-O-Si bond-angle distribution, $V(\alpha)$. For $\text{K}_2\text{Si}_4\text{O}_9$ glass this value is $\eta = 0.20$, which corresponds to a bond angle of 143° . To derive the entire bond-angle distribution we must transform the distribution in asymmetry parameters, $X(\eta)$, obtained from the 2-D NMR experiment to a distribution of bond angles, $V(\alpha)$, using the following equation⁹

$$V(\alpha) = X(\eta) |d\eta/d\alpha| \quad (1)$$

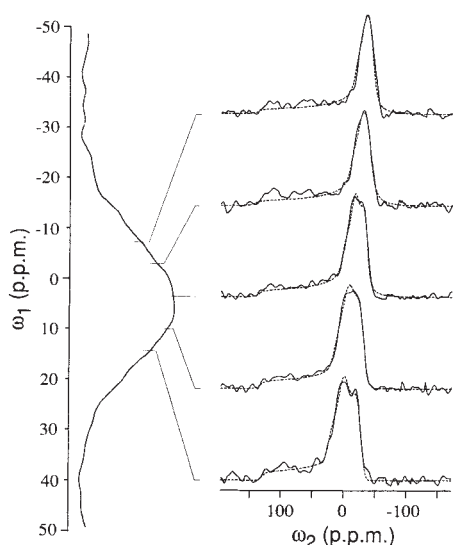


FIG. 2 Details of the bridging oxygen ^{17}O NMR resonance in $\text{K}_2\text{Si}_4\text{O}_9$ glass. Slices parallel to the ω_2 anisotropic axis (solid lines) are shown together with the results of multi-parameter least-squares fits (dotted lines) to obtain the variation of the magnitude and asymmetry of the electric field gradient across the isotropic line shape (ω_1). The residual chemical shift anisotropy and dipolar coupling present in the ω_2 line shape were approximated by a line-broadening parameter.

A partial Si-O-Si bond angle distribution, $V(\alpha)$, determined solely from our fitted values of η and the correlation of η with bond angle is shown by the black dots in Fig. 4. But the magnitude of the oxygen EFG also varies systematically with bond angle, according to theoretical calculations²³ and our own data. The magnitude of the oxygen EFG closely follows an empirical expression for the variation in oxygen sp hybridization with Si-O-Si angle²⁴, $(\cos \alpha / (\cos \alpha - 1))$. Values of the EFG obtained from fits of ω_2 line shapes are plotted against the angle derived from the value of η obtained from the same fit in Fig. 3b. Also shown are the results of earlier molecular orbital calculations of oxygen EFG²³, where the absolute magnitude has been scaled to coincide with the experimental curve. This relationship between oxygen EFG and α may also be used to determine $V(\alpha)$ simply from the line shape in ω_1 . The isotropic quadrupolar shift, $\delta_{\text{iso}}(Q)$ in equation (2) (where ν_L is the Larmor frequency and I the spin quantum number), obtained in ω_1 by subtracting the chemical shift from the total shift

$$\delta_{\text{iso}}(Q) = -\frac{3}{40}[(I(I+1) - \frac{3}{4})/I^2(2I-1)^2](P_q/\nu_L^2) \times 10^6 \quad (2)$$

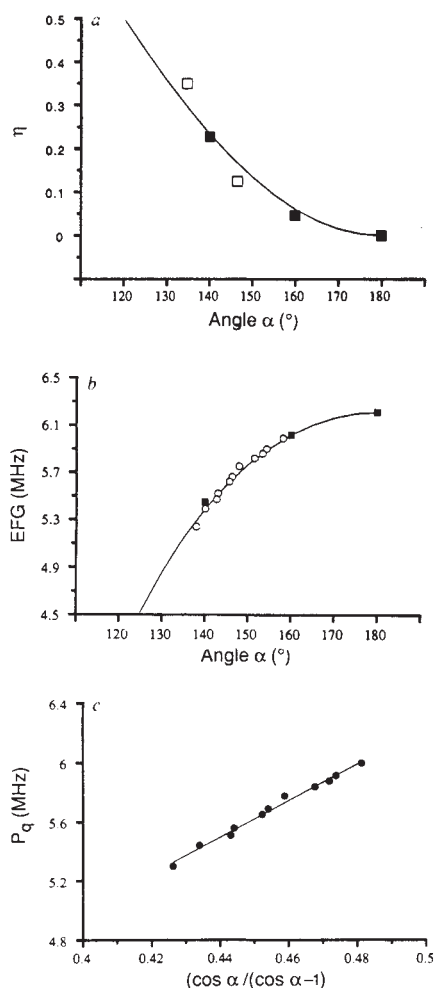


FIG. 3 Electric field gradient (EFG) parameters for bridging oxygens as a function of the Si-O-Si bond angle (α). a, Asymmetry parameter η . (■) calculated by Tossell and Lazzeretti²³, (□) experimental points for bridging oxygens in cristobalite ($\alpha = 146.4^\circ$)³³ and wadeite¹⁵ ($\alpha = 134.7^\circ$). The solid line is the function $\eta = 1 + \cos \alpha$. b, Magnitude of EFG (○) for the fitted slices parallel to ω_2 in the $\text{K}_2\text{Si}_4\text{O}_9$ glass spectrum with the angle given by the fitted value of η . ■, The trend in calculated values from ref. 23 scaled to lie on the experimentally determined line. The solid line is the relationship $\text{EFG}(\alpha) = \text{EFG}(180^\circ) \times 2(\cos \alpha / (\cos \alpha - 1))$. c, Quadrupolar product P_q [$\text{EFG} \times (1 + \frac{1}{3}\eta^2)^{1/2}$] determined from fits of ω_2 line shapes against $(\cos \alpha / (\cos \alpha - 1))$. The straight line has a correlation coefficient of $r = 0.995$.

depends on the quadrupolar product (P_q) of the magnitude of the oxygen EFG and its asymmetry, η

$$P_q = \text{EFG} \times (1 + \frac{1}{3}\eta^2)^{1/2} \quad (3)$$

Because $(1 + \frac{1}{3}\eta^2)^{1/2}$ is close to 1 ($0 \leq \eta \leq 1$), P_q shows a similar correlation with $(\cos \alpha / (\cos \alpha - 1))$ as the EFG (Fig. 3c). Consequently, $V(\alpha)$ can be obtained by two transformations, using equations similar to equation (1), from a distribution in $\delta_{\text{iso}}(Q)$ to a distribution in P_q and then to a distribution in α . This is the full distribution shown in Fig. 4.

The experimentally derived distribution $V(\alpha)$ for $\text{K}_2\text{Si}_4\text{O}_9$ glass has a peak at 143° and half-width of 21° from 135° to 156° . Mozzi and Warren⁷ have derived an inter-tetrahedral bond-angle distribution for SiO_2 glass which is shown for comparison in Fig. 5. The $\text{K}_2\text{Si}_4\text{O}_9$ glass distribution is narrower but follows the shape of the Mozzi and Warren distribution with the peaks coinciding (144 and 143°) and both being skewed towards low angles and tailing off towards high angles. Both distributions probably reflect repulsion between Si atoms at low bond angles and a shallow dependence of energy on angle out to 180° (ref. 25).

The main difference between SiO_2 and $\text{K}_2\text{Si}_4\text{O}_9$ is the presence of nonbridging oxygens. Apart from the static structural effect introduced by the network modifier, the nonbridging oxygens created also provide a flow mechanism for the silicate liquid²⁶ which allows structural relaxation to occur more readily. In $\text{K}_2\text{Si}_4\text{O}_9$ glass, it seems that more Si-O-Si bonds are able to take up the energetically most favourable angle causing a narrowing of the total distribution. The flow mechanism involves a change in oxygen speciation from nonbridging to bridging (or vice versa) by means of the approach of a nonbridging oxygen to a silicate tetrahedron and the formation of a silicon intermediate with fivefold coordination²⁷. It seems that when a nonbridging oxygen finds itself in an ordered arrangement with network modifiers, such as that observed by ^{17}O NMR in the glass, it becomes trapped in this low-energy configuration. It is then unavailable to take part in the structural relaxation mechanism. Presumably, as this happens throughout the liquid, structural relaxation is progressively inhibited and the glass transition occurs. The effect of the strength of the modifier oxygen bonds and the glass transition temperature on silicate glass structure

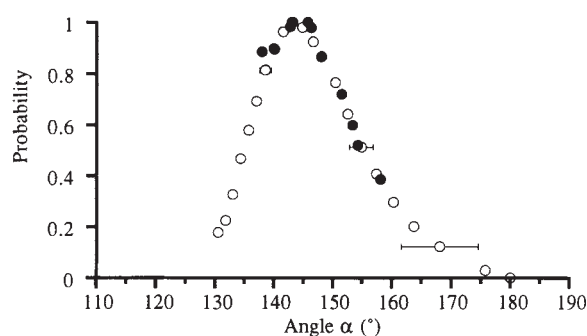


FIG. 4 Si-O-Si bond-angle distributions ($V(\alpha)$) for $\text{K}_2\text{Si}_4\text{O}_9$ glass (●) determined from fitted values of η . Less-probable bond angles have lower signal intensity and are difficult to fit, so only a partial bond-angle distribution is shown. (○) Determined from the relationship between P_q and Si-O-Si bond angle. P_q is determined from the isotropic quadrupolar shift in ω_1 ($\delta_{\text{iso}}(Q)$, equation (2)). Values of $\delta_{\text{iso}}(Q)$ beyond those obtained by fitting slices in ω_2 were separated from the chemical shift by extrapolation of a linear relationship between total shift and chemical shift for the fitted slices. Error bars on the distribution determined from P_q are propagated from the error in determining the isotropic quadrupolar shift and hence P_q . This results in greater uncertainty in the high bond angles because P_q is slowly varying as it approaches 180° .

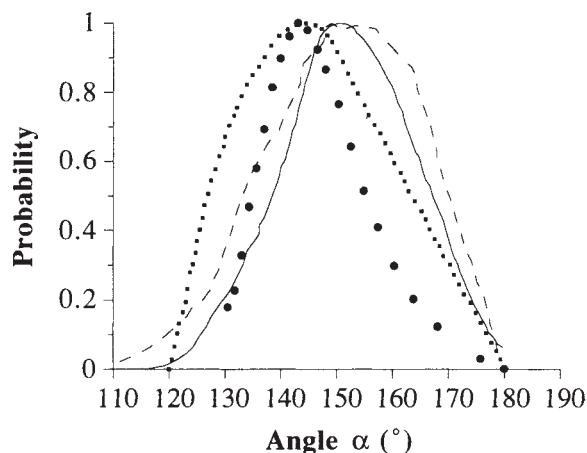


FIG. 5 Comparison of bond-angle distributions ($V(\alpha)$) for silicate glasses: ●, $K_2Si_4O_9$ reported here; ■, SiO_2 determined from the X-ray radial distribution function by Mozzi and Warren⁷. (---) $78K_2O \cdot 216SiO_2$ produced by a molecular dynamics simulation with a two-body potential²⁹; (—) $Na_2O \cdot 2SiO_2$ produced by a molecular dynamics simulation using a three-body potential³¹.

observed by means of the Si-O-Si bond-angle distribution should allow detailed investigation of these processes.

The bond-angle distribution for $K_2Si_4O_9$ glass does not show the bimodal bond-angle distribution with an additional peak at $\sim 110^\circ$ that was present in earlier molecular dynamics simulations of modified silicate glasses²⁸. More-recent simulations of potassium silicate glasses^{29,30} produce a bond-angle distribution that is opposite in overall shape (skewed to high angles) and much broader than the distribution here (Fig. 5). The broadness is certainly a result of the high quench rates (10^{14} K s^{-1}) and therefore high glass transition temperature of these simulated glasses. The opposite skewing probably reflects the nondirectional nature of the two-body potentials used. They cannot reproduce the repulsive effects that cause the rapid fall-off of the distribution to zero around 120° . Molecular dynamics simulations of sodium silicate glasses³¹ employing three-body potentials are in better agreement but are still skewed towards high angle. Realistic quench rates, and therefore distribution widths, are unlikely in these simulations because of computer limitations, but overall reproduction of the shape of the distribution should be a good test of the potentials being used.

Prospects for future work

The ability to characterize the inter-tetrahedral bond-angle dis-

tribution should provide insight into the variation of glass structure with pressure and glass transition temperature and test the accuracy of model distributions derived from other experimental techniques and computer simulations. Because the technique probes the bridging angle locally, complex compositions do not cause problems as they do in scattering experiments where there are many overlapping pair correlation functions in the radial distribution analysis. Silicon-29 NMR has also been used to derive Si-O-Si bond-angle distributions for SiO_2 glass^{9,10,32} but it measures the average of four such angles at each SiO_4 tetrahedron and uses a correlation that does not hold if a tetrahedron contains a nonbridging oxygen. The distribution derived here is sensitive to each Si-O-Si bond and is not affected by nonbridging oxygens unless by coincidence they overlap the bridging oxygen resonance.

In general, similar 2-D NMR techniques could be applied to any material in which there are distributions of environments causing a broad NMR line in one dimension. By resolving the data into isotropic and anisotropic dimensions, we should find the additional symmetry and population information about site distribution very useful in characterizing sites in catalysts, amorphous semiconductors and incommensurate materials as well as glasses. □

Received 2 March; accepted 2 June 1992.

- Greaves, G. N., Fontaine, A., Lagarde, P., Raoux, D. & Gurman, S. J. *Nature* **293**, 611-616 (1981).
- Greaves, G. N. in *Glass Science and Technology* (eds Uhlman, D. R. & Kreidl, N. J.) Vol. 4B, 1-76 (1990).
- Eckersley, M. C., Gaskell, P. H., Barnes, A. C. & Chieux, P. *Nature* **335**, 525-527 (1988).
- Gaskell, P. H., Eckersley, M. C., Barnes, A. C. & Chieux, P. *Nature* **350**, 675-677 (1991).
- Dupree, R., Holland, D., McMillan, P. W. & Pettifer, R. F. *J. non-cryst. Solids* **68**, 399-410 (1984).
- Stebbins, J. F. *J. non-cryst. Solids* **106**, 359-369 (1988).
- Mozzi, R. L. & Warren, B. E. *J. appl. Cryst.* **2**, 164-172 (1969).
- Dupree, R. & Pettifer, R. F. *Nature* **308**, 523-525 (1984).
- Pettifer, R. F., Dupree, R., Farnan, I. & Sternberg, U. *J. non-cryst. Solids* **106**, 408-412 (1988).
- Geissberger, A. E. & Bray, P. J. *J. non-cryst. Solids* **54**, 121-137 (1983).
- Kirkpatrick, R. J. et al. in *Structure and Bonding in Non-Crystalline Solids* (eds Walfraen, G. E. & Revesz, A. G.) (Plenum, New York, 1986).
- Samoson, A., Lipmaa, E. & Pines, A. *Molec. Phys.* **65**, 1013-1018 (1988).
- Chmelka, B. F. et al. *Nature* **339**, 42-43 (1989).
- Lior, A. & Virlet, J. *Chem. Phys. Lett.* **152**, 248-253 (1988).
- Mueller, K. T. et al. *J. magn. Reson.* **86**, 470-487 (1990).
- Eastman, M. A., Grandinetti, P. J., Lee, Y. K. & Pines, A. *J. magn. Reson.* (in the press).
- Cohen, M. H. & Reif, F. *Solid State Phys.* **5**, 321-438 (1957).

- Mueller, K. T., Baltisberger, J. H., Wooten, E. W. & Pines, A. *J. phys. Chem.* (in the press).
- Mueller, K. T., Wu, Y., Chmelka, B. F., Stebbins, J. F. & Pines, A. *J. Am. chem. Soc.* **113**, 32-38 (1991).
- Xue, X., Stebbins, J. F. & Kanzaki, M. *Trans. Am. Geophys. Un.* **72**, 572 (1991).
- Schweinsberg, H. & Liebau, F. *Acta cryst.* **B30**, 2206-2213 (1974).
- Vessal, B. et al. *Nature* **356**, 504-506 (1992).
- Tossell, J. A. & Lazzeretti, P. *Phys. Chem. Miner.* **15**, 564-569 (1988).
- Engelhardt, G. & Radeaglia, R. *Chem. Phys. Lett.* **108**, 271-274 (1984).
- Newton, M. D. & Gibbs, G. V. *Phys. Chem. Miner.* **6**, 221-246 (1980).
- Farnan, I. & Stebbins, J. F. *J. Am. chem. Soc.* **112**, 32-39 (1990).
- Stebbins, J. F. *Nature* **351**, 638-639 (1991).
- Soules, T. E. *J. chem. Phys.* **71**, 4570-4578 (1979).
- Huang, C. & Cormack, A. N. *J. chem. Phys.* **95**, 3634-3642 (1991).
- Tesar, A. A. & Varshneya, A. K. *J. chem. Phys.* **87**, 2986-2989 (1987).
- Melman, H. & Garofalini, S. H. *J. non-cryst. Solids* **134**, 107-115 (1991).
- Gladden, L. F., Carpenter, T. A. & Elliott, S. R. *Phil. Mag.* **B53**, L81-L87 (1986).
- Spearing, D. R., Farnan, I. & Stebbins, J. F. *Phys. Chem. Miner.* (in the press).

ACKNOWLEDGEMENTS. This investigation was supported by the Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division of the U.S. Department of Energy (A.P.) and grants from U.S. NSF, Earth Sciences Division (J.F.S.), P.J.G. and M.A.E. acknowledge support from the NIH, J.H.B. from the NSF, and U.W. from the Alexander von Humboldt Foundation.

Molecular diversity of the NMDA receptor channel

Tatsuya Kutsuwada^{*†}, Nobuko Kashiwabuchi^{*}, Hisashi Mori^{*}, Kenji Sakimura^{*}, Etsuko Kushiya^{*}, Kazuaki Araki^{*}, Hiroyuki Meguro^{*†}, Hisashi Masaki^{*}, Toshiro Kumanishi[‡], Masaaki Arakawa[†] & Masayoshi Mishina^{*§}

Departments of ^{*}Neuropharmacology and [‡]Neuropathology, Brain Research Institute and [†]Department of Medicine II, School of Medicine, Niigata University, Niigata 951, Japan

Two novel subunits of the mouse NMDA receptor channel, the $\epsilon 2$ and $\epsilon 3$ subunits, have been identified by cloning and expression of complementary DNAs. The heteromeric $\epsilon 1/\zeta 1$, $\epsilon 2/\zeta 1$ and $\epsilon 3/\zeta 1$ NMDA receptor channels exhibit distinct functional properties in affinities for agonists and sensitivities to competitive antagonists and Mg^{2+} block. In contrast to the wide distribution of the $\epsilon 1$ and $\zeta 1$ subunit messenger RNAs in the brain, the $\epsilon 2$ subunit mRNA is expressed only in the forebrain and the $\epsilon 3$ subunit mRNA is found predominantly in the cerebellum. The $\epsilon 1/\zeta 1$ and $\epsilon 2/\zeta 1$ channels expressed in *Xenopus* oocytes, but not the $\epsilon 3/\zeta 1$ channel, are activated by treatment with 12-*O*-tetradecanoylphorbol 13-acetate. These findings suggest that the molecular diversity of the ϵ subunit family underlies the functional heterogeneity of the NMDA receptor channel.

THE glutamate receptor (GluR) channel plays a key role in brain function. Most of the fast excitatory synaptic transmission is mediated by GluR channels in the central nervous system¹. Furthermore, GluR channels are essential for synaptic plasticity such as long-term potentiation and long-term depression, which are thought to underlie memory acquisition and learning^{2,3}. Involvement of GluR channels in network formation and degenerative brain disorders has been also suggested⁴⁻⁶. On the basis of the pharmacological and electrophysiological properties, GluR channels have been classified into three major subtypes that are the α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA), kainate and *N*-methyl-D-aspartate (NMDA) receptors^{1,7}.

Numbers of mammalian GluR channel subunits have been characterized by cloning and expression of the cDNAs⁸⁻²². These GluR channel subunits possess four putative transmembrane segments characteristic of neurotransmitter-gated ion channels and have been classified into several subfamilies according to the amino-acid sequence homology¹³⁻²¹ (see Table 1). The members of the α subfamily constitute AMPA-selective GluR channels⁸⁻¹², whereas the β and γ subfamilies include the subunits of the kainate-selective GluR channel¹⁴⁻¹⁷. Expression of the NMDAR1 ($\zeta 1$) subunit as well as the $\zeta 1$ -2 subunit, a smaller form of the $\zeta 1$ subunit, can produce NMDA receptor channels in *Xenopus* oocytes^{18,19}. But highly active NMDA receptor channels are formed only when the $\epsilon 1$ and $\zeta 1$ subunits are expressed together²¹. The ϵ and ζ subfamilies thus represent the subunits of the NMDA receptor channel.

The NMDA receptor channel is highly permeable to Ca^{2+} and is essential for synaptic transmission and synaptic plasticity underlying memory, learning and development^{1,2,4,23}. In view of the diverse functions of the NMDA receptor channel, it is possible that the NMDA receptor channel exists in multiple forms with different properties. In fact, the regional variation in ligand binding properties of the NMDA receptor has been observed in the central nervous system^{24,25}. In the present investigation, the primary structures of two novel mouse NMDA receptor channel subunits, designated as $\epsilon 2$ and $\epsilon 3$, have been revealed by cloning and sequencing the cDNAs. Expression from the cloned cDNA of the $\epsilon 2$ or $\epsilon 3$ subunit together with

the $\zeta 1$ subunit in *Xenopus* oocytes yields functional NMDA receptor channels. Three ϵ subunits are distinct from each other in pharmacological properties and distribution in the brain. Furthermore, modulation of the channel activity by 12-*O*-tetradecanoylphorbol 13-acetate (TPA) treatment is described.

Molecular diversity

Two novel members of the mouse NMDA receptor channel, designated as the $\epsilon 2$ and $\epsilon 3$ subunits, were identified by screening of a mouse brain cDNA library with an $\epsilon 1$ subunit cDNA as a probe under low stringency conditions. The entire primary structures of the $\epsilon 2$ and $\epsilon 3$ subunits were deduced using several overlapping cDNA clones (Fig. 1). The initiator methionine is assigned to the first methionine residue found in a large open reading frame. Analysis and comparison of the local hydrophobicity profiles suggest the presence of an amino-terminal signal peptide and four putative transmembrane segments (M1-M4) in the middle of the molecule. The proposed mature $\epsilon 2$ and $\epsilon 3$ subunits are composed of 1,456 and 1,220 amino acids with calculated molecular masses of 162,875 and 133,513, respectively. The $\epsilon 2$ and $\epsilon 3$ subunits share high amino-acid sequence identity with the $\epsilon 1$ subunit and thus belong to the ϵ subfamily of GluR channel subunits (Table 1). The sequence identity with the other rodent GluR channel subunits thus far reported is 12–18%. As is the case for the $\epsilon 1$ and $\zeta 1$ (NMDAR1) subunits of the NMDA receptor channel^{18,19,21}, the $\epsilon 2$ and $\epsilon 3$ subunits possess an asparagine residue (amino-acid residues 589 and 593, respectively) at the position corresponding to the arginine residue 586 in M2 segment of the $\alpha 2$ subunit which determines the Ca^{2+} permeability of the AMPA-selective GluR channel²⁶⁻²⁸. This asparagine residue conserved among NMDA receptor channel subunits may play a role in high Ca^{2+} permeability and Mg^{2+} blocking of the channel.

Functional diversity

The $\epsilon 2$ or $\epsilon 3$ subunit-specific mRNA was synthesized *in vitro* from cloned cDNA and was injected into *Xenopus* oocytes together with the $\zeta 1$ subunit-specific mRNA. The peak inward currents obtained in normal frog Ringer's solution at -70 mV membrane potential were 667 ± 187 nA (mean $\pm$ s.e.m.) and 546 ± 172 nA ($n = 5$) in response to $10 \mu M$ L-glutamate plus $10 \mu M$ glycine and $100 \mu M$ NMDA plus $10 \mu M$ glycine, respec-

§ To whom correspondence should be addressed.

FIG. 1 Alignment of the deduced amino-acid sequences (single-letter code) of the $\epsilon 1$ (ref. 21), $\epsilon 2$, $\epsilon 3$ and $\zeta 1$ (ref. 19) subunits of the mouse NMDA receptor channel. Amino-acid residues are numbered beginning with the N-terminal residue of the proposed mature subunit (indicated by arrowheads)³⁶; the cleavage site assigned for the $\epsilon 1$ subunit is different from the previous one²¹. Numbers of the amino-acid residues at the end of the individual lines are given. Sets of identical amino-acid residues among four subunits or three ϵ subunits are boxed. The putative transmembrane segments (M1–M4) are indicated. Potential phosphorylation sites for Ca^{2+} -calmodulin-dependent protein kinase type II (ref. 37) are marked with open circles. The nucleotide sequence data will appear in the DDBJ, EMBL and GeneBank Nucleotide Sequence Database under the accession numbers D10651 and D10694.

METHODS. Screening of a mouse brain cDNA library under low-stringency conditions with the 1,383 base-pair (bp) *KpnI*–*HindIII* fragment from the plasmid pGRA19 (ref. 21) as a probe yielded one $\epsilon 2$ subunit cDNA clone U9 and two $\epsilon 3$ subunit cDNA clones U20 and He49. The cDNA inserts were subcloned into the *EcoRI* site of the plasmid pBluescript II SK(–) (Stratagene) to yield the plasmids pGRU9, pGRU20 and pGRHe49. Additional clones were isolated by screening under high-stringency conditions using the 445 bp *DraIII*–*EcoRI* and 2,847 bp *HindIII*–*EcoRI* fragments from pGRU9, the 1,615 bp *EcoRI* fragment from pGRHe49 or the 314 bp *HindIII*–*EcoRI* fragment from pGRUN3 (see below) as probes. The cDNA inserts of these recombinant phage were subcloned into the *EcoRI* site of the plasmid pBKSa (ref. 19). The complete nucleotide sequences of the $\epsilon 2$ subunit cDNA clones U9 (nucleotide residues –417–3,218) and U4 (1,117–4,454) were determined on both strands by the dideoxy chain termination method³⁸. Partial nucleotide sequences of the clones U2 (residues 4,319–4,396), U7 (–417–3,616), U8 (4,319–4,396), U11 (4,319–4,396), U16 (–205–57), U17 (–205–57) and U22 (1,449–4,365) are in complete agreement with those of the clones U9 and U4. The nucleotide sequences of the $\epsilon 3$ subunit cDNA clones UN6 (–232–3,283),

$\epsilon 1$	MGR LGY TLL LUL PAL LUH G P A A A A E K G T P A L N I A U L L G H S H D U T E R L N L W G P E A T G P L D U N U A L L N N A T O P K S L I T H U C D L S G A R I H A L U	77
$\epsilon 2$	MKPSAECCSPKFWLULALAVUSGSKARSKSAPSIGIAU LUGTSD--EVAIKDAHEKDDFHLSUUPRVELUAMNETOPKSIITRICOLMSDRKIQOMU	72
$\epsilon 3$	MGGALGPALLTSL LG A H A G L G A G G E A U T A U U F G S S G P L - Q A A Q A R T L T P N F L D L P L E I O P L T I G U N T N I S S I L T O I C G L L G A A R U H G I U	76
$\zeta 1$	MSTHLLT F A L L F S C S F A A R A C D P K I U N I G A U L S T R K H E Q N F R E A U N Q A N K R H S W K I Q L N A T S U T H K P N A I G H A L S U C E D L I S S Q U A I L U S	75
$\epsilon 1$	F O D D T D E A V A Q M L D F I S S Q T F I P I L O H G A S M I A M K D P T I S T F F G A S I Q Q A T I M L K I M Q Y D M H F S L U T I F P Y R D F I S F I K T U D S F U G D	177
$\epsilon 2$	L A N D T D E A I A Q I L D F I S A Q T I P I L G I H G S S M I A M K D E S T F F G A S I Q Q A S L I N I E E Y D M Y I F I S I U T I Y F P Q Y D F U N K I R S T I E N S F U G E	172
$\epsilon 3$	F E N U D T E A V Q L D F V S S O T H I P I L S I S G S A U L T P K E P S A F L Q L G S L E Q L Q A L F K U L E E Y D S A F A U I S L M P G A L E G U R A A D A S Y L S R	175
$\zeta 1$	H P T P N D H F T P T U S Y T A G F Y R I P M L G L - T T R S I Y S D S I H S L R T U P P Y S S S S F E M H R U Y M N H I I L L U S D D H E G R A - A A K R L E T L E E R E S K	172
$\epsilon 1$	M A N D I T D T S F E D - - A K T Q U L K K I H S S U I L L Y S K D E A U L I L S E A R S L G T I Y D F F I U P S L U S A N T E L I P K E F P S L I S U S Y D D M Y S L E A R D R D L G	275
$\epsilon 2$	L E E V L L D M S L D D G S K I O N L K K L G S P I L L Y C T K E A Y I F E A N S U G L I G Y G T M I U P S U A G D I D T U P S E F T I G L I S U S Y D E D Y G L P A R U R D G I A	272
$\epsilon 3$	L L D U L T L E L G P G P R A R T O R L R Q U D A P U L U A Y C S R E A E U L F A E M A Q A G L V O P G H U L U P N L A L S T D A P P A A P M G L I S U T E S M R L S L R Q K U R D G V A	275
$\zeta 1$	A E K L G F D P G T K N U T A L L M E A - R D L E A R U I L S A S E D D A T U Y R A M L N T I G S Y U L U G E - - R E I S G N A L Y A D D I G L Q L I N K N E S A H I S D A U G U	289
$\epsilon 1$	I L T T A S S M L E K F S Y I P E A K A S Y G G T E K P E T P L H T L H O F M U N U T U A K D I S T E E G Y U H P R L U U L U L N K D E K U G K L E N O T L - R L R H A U P - R Y K S	373
$\epsilon 2$	I I T T A S D M L S E N S F I P E K P S S Y N T H E K R I Y Q S N M L N R Y L I N U T F E G R N L S F S E G Y A M H P K L U I I L N K E K K E R U G K L D K S L - O M K Y Y U M P - R M C P	378
$\epsilon 3$	I L A G A S Y R R Q Y G T L P A P A G D C R S H P G P S P A R E A F Y R H L L N U T W E A R D F S F S P G G Y L U O P T H U M L N R H L E W E U G R A D H G U L L - Y M K Y P U M P - R Y S T	373
$\zeta 1$	U A Q A U H E L L E K E N I T D P R G C U G N T I N W K T G P L F K R L N S S K Y A D G U T G R U E F N E D G R K F A N Y S I M L N R K L U G U G I Y N G T H U I P N D R K I I U G G E T E	369
$\epsilon 1$	F S D C E P D D M H S I U T L E E A P F U I U - - - E D I D P L T E T U R N T U - P C R K F U K I N N S T N E G N - U K C C K G F C I D I L K K L S R T U K F I Y D L Y L U T N G K - - -	462
$\epsilon 2$	E T E - E Q E D H S I U T L E E A P F U I U - - - E S U D P L S G T C R N T U - P C Q K R I S E N K T D E E P Y I K C C K G F C I D I L K K I S K S U K F I Y D L Y L U T N G K - - -	459
$\epsilon 3$	S L O P U D S R H L T U A T L E E A P F U I U - - - E S P D P G T G G C U N T U - P C R R Q S N H T F S S G D I T P Y K L C C K G F C I D I L K K L A K U S Y D L Y L U T N G K - - -	463
$\zeta 1$	K P R G Y Q H S T R L K I U T I H Q E F U I U K P T H S D G T C K E E F T U N D P K K I C T G P N D T S P G S P R H T U P C C Y G F C D I L L L A R T N T I T E U H L U A D G K F O T Q	489
$\epsilon 1$	- - H O K K U N N U M G H I G E U Y U A U G S L T I N E E R S E U D F S U P F I E T G I S U H S R S N G T U S P S A F L E P S A U U M H F U L I U S A I A U F E Y F S P U	568
$\epsilon 2$	- - H G K I N G T W N G H I G E U M K R A M A U G S L T I N E E R S E U D F S U P F I E T G I S U H S R S N G T U S P S A F L E P S A U U M H F U L I U S A U A U F E Y F S P U	557
$\epsilon 3$	- - H G K R U G U W N G H I G E U Y K R A M A U G S L T I N E E R S E I I D F S U P F I E T G I S U M H S R S N G T U S P S A F L E P S A U U M H F U L I U S A I A U F E Y F S P U	561
$\zeta 1$	E R U N S N K K E W N G H I D E L L S G Q D M I U A P L T I N E R A Y I E S K P F K Y Q L T I L K K E I P R S T L D S F M O D Q S T L L U - G L S U H U M A U N L Y L D R E S P	567
$\epsilon 1$	G Y N R N L A K A K A P H S F T I G K A I W L L G L U F N S S U P U N P K G T T S K I M U S U A F F A U I F L A S Y A N L A A F M I D E E F U D O U I G L S D K K F Q R P H Y S P P F R F	668
$\epsilon 2$	G Y N R C L A D R E P G P S F T I G K A I W L L G L U F N S S U P U N P K G T T S K I M U S U A F F A U I F L A S Y A N L A A F M I D E E Y U D O U I G L S D K K F Q R P D F S P P F R F	657
$\epsilon 3$	S Y N N L I T K K K S G P S F T I G K S U W L L U A F N S S U P I E N P R A T T S K I M U L U A F F A U I F L A S Y A N L A A F M I D E E Y I D T U S G L S D K K F Q R P Q D Y P P F R F	661
$\zeta 1$	- F G R F K U N S E E E E D A L T S S A M F S U G U L L N S G I G E A R S F S A R I L G H U A F A U I U A S Y A N L A A F L U L D R P E E R I T G I N D P R L N E - - - S D K F I Y	663
$\epsilon 1$	G U P N G S T E R N I R N N - Y P Y M H A Y M K F N Q R Q U D A L S L K T G K L D A F I Y D A U L N Y A G R D E G C K L U T I G S G V I F A T I G Y G I A U K G S W K R Q I D L A L L	758
$\epsilon 2$	G U P N G S T E R N I R N N - Y A E M H A Y M K F N Q R Q U D A L S L K T G K L D A F I Y D A U L N Y A G R D E G C K L U T I G S G V I F A T I G Y G I A U K G S W K R Q I D L A L L	755
$\epsilon 3$	G U P N G S T E R N I R N N - Y R D M H A Y M K F N Q R Q U D A L S L K T G K L D A F I Y D A U L N Y A G R D E G C K L U T I G S G V I F A T I G Y G I A U K G S W K R Q I D L A L L	759
$\zeta 1$	A T L Q S S U D I Y F R Q U E L S T Y R N H E K M Y E S A E E I Q A U R D N K L A F I D S A U L E F E S - - Q R C L U T - T G L L F R S G A G H R K D S P K O N U S I L	759
$\epsilon 1$	Q F U D G E E E L E T L W L G I C H N E K N E U M S S O L D I D N N A G U F Y M L A A T A L S I T I F I C H L F Y W K L R F C T G U C S D R P G L L F S I S R G I Y S C I H G U H I E E K	858
$\epsilon 2$	Q L F D G E E E L E A L W L G I C H N E K N E U M S S O L D I D N N A G U F Y M L A A T A L S I T I F I C H L F Y W K L R F C H G U C S G K P G H U F S I S R G I Y S C I H G U A I E E R Q	855
$\epsilon 3$	Q L F D G E T O K L E T U N S G I C H N E K N E U M S S O L D I D N N A G U F Y M L A A T A L S I T I F I C H L F Y W K L R H S U P S - S S A L D F R S G I Y S C I H G U S L P S P	857
$\zeta 1$	K S H E N G F M E D L K T M U R Y Q E C D S R S N A - P A T I F E N N A G U F L U A G G I V A G I F I F I E I A Y K R H K D A R R K Q M L A F A A U N W R K N L Q D R K S G R A E P D P K K	858
$\epsilon 1$	- - - K S P D F N L T G S S N H L K L R S A K N I S N H S N H S R M D S P K R A D F I Q S L I U D H S D K N L I Y S D N R S F Q G - - - K D S I G E N N H L Q - T F U A N R H K D	951
$\epsilon 2$	S U M S P I A T N N H S N L R L R T A K N M A N L S G U N - - - G S P Q S A L D F I R R E S S Y D I S E H R R S F T H S D C K S Y N N P P C E E N L F S D Y I S E V E R T F G N L Q L K D	951
$\epsilon 3$	A R P P S P L T A G S A Q A N Y L K L Q A A R D M U S T A - D U S G S L D R A T T I E N W G N N R A P A T T S G P R S C T P G P G O P S G W R P P G G O R T P L A R R A P Q P P A R P G	956
$\zeta 1$	K A T F R A I T S L A S S F K R R R S S K D I S T G G R G A L Q N Q D T U L P R A I E R E E G L Q L C S R H R E S	920
$\epsilon 1$	S L S N Y U F Q Q H P L T L N E S N P M T U E V A U S T E S K N S R P R L W K K S H S L R O D S L N Q - P U S Q R D E K T A E N R T H S L K S P R Y L - P E E V A H S D I S E T S S R A T C H	1049
$\epsilon 2$	S - - - N U V Q D Y H - - - H N H R P H S I G S T S S I D G L Y D C D N P F T T Q P R - S I S K P L D I G L P S S K H S A L D Y G K F S K S D R Y S G H D L I R S D U S D I S T H U T Y	1044
$\epsilon 3$	P A Q O R L S P T C P E H P A G T L O N R O G Q C E S G I R D T S R P P E R A L P E R S L L H A C H Y S S F P R A E R S A R P F L P F P E P P E D D L P L L G P E Q L A R R E A L R A W A	1050
$\epsilon 1$	R E P D N N K N H K T D N F - - - K R S H A S K Y P K D C S E V E R T Y U K T K A S S P R D K I Y T I D G E K P S F H L D P P Q F I E N I U L P E N U D F P D T Y Q D H N E F R K G D S T L P	1144
$\epsilon 2$	G N I E G N A K R R K Q Y K D S L K R P A S A K S R R E F D E I E L A Y R R P P R S P D H K Y F R D K E G L R D F Y L D F R T K E N S P H W E H U D L T I H K S C D F K R - D S U G	1143
$\epsilon 3$	R G P R P R H S L P S S U A E A F T R S N L P A R C T G H A C A C P Q S R P S R C H U A G T Q S L R L P S Y E A C U E G U P A G U A T W Q P R Q H U C L H T H L P F C W G T U C R H P P	1156
$\epsilon 1$	M H R N P L H N E D G L P N N D Q Y K L A K H F T L K D G S P H S E G S D R Y R Q N S T H C R S C L S N L P T Y - - - S G H F T H R S P - F K C D A C L R M G N L Y D I D E D Q L Q E T G N P A	1239
$\epsilon 2$	G O P C T N R S H L K H T G D K H G U U G G U P A P W E K N L T U D W E D - - - S G N F C R S C P S K L H N Y S T U A G Q N S G R A C I R C A C K K A G N L Y D I S E D N S L Q E L D Q P A	1241
$\epsilon 3$	P C S S H S P L I G T W E P S H R G R T L G L T G Y R D S G U L E U S R E A C T G G F P R S C T W R I S S L E S E U	1220
$\epsilon 1$	T R - E - - E A Y Q Q D W S Q N - - N A L Q F Q K N K L I N R Q H S Y D N I L D K P R I D L S R P S R S I S L K D R E R L L E G N L Y S L F S P U S P - K L L G N K S S L F P Q G L E S K - -	1338
$\epsilon 2$	A P U A U S S M A S T T K Y P Q S P T N S K A Q K K N R N L R Q H S Y D T F U D L Q E - E A L A P R S U S L K D K R F N D G S P Y A H N F E H P A G E S S F A N K S S U T A G H H N N P G	1340
$\epsilon 1$	- - - R S K S L L P D H T S D N P L F L T Y G D D Q R L U I G R C P S D P Y K H S L P S Q A U N D S Y L R S S L A S T A S Y C - - - S R D S R G H S D U Y I S E H U M P Y A A N K N N H Y S T P R U	1422
$\epsilon 2$	S G Y M L S K S L Y P D R U T Q N P I P T F G D D Q C L L A N P T S S G S P T U A G A S K T R P D F R A L U T N K P U S A L H G A U P G R F Q K D I C I G N Q S N P C U P N N K N - - - P R A	1436
$\epsilon 1$	L N S C S N R R U Y K M P S I E S D U	1442
$\epsilon 2$	F N G S S N G H U E K L S S I E S D U	1456

He49 (–101–1,504), UN3 (52–4,039), U20 (347–2,016) and UT5 (2,487–4,071) are in agreement except for the residues 711 (T or C), 1,485 (A or G), 1,494 (A or G) and 2,721 (T or C). Nucleotide residues are numbered in the 5' to 3' direction, beginning with the codon specifying the N-terminal residue of the proposed mature subunit. Analysis of nucleotide and amino-acid sequences was done using GENETYX software (SDC) unless otherwise specified.

TABLE 1 Per cent amino-acid sequence identity between rodent GluR channel subunits

Subfamily	α				β		γ		δ	ϵ			ζ
Subunit	$\alpha 1$	$\alpha 2$	GluRC	GluRD	GluR5-1	$\beta 2$	KA-1	$\gamma 2$	$\delta 1$	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\zeta 1$
$\alpha 1$		66	64	66	36	37	33	32	25	14	13	15	18
$\alpha 2$			72	70	36	37	31	30	23	13	13	15	20
GluRC				71	36	38	32	31	22	12	12	14	19
GluRD					36	36	31	30	23	13	13	15	20
GluR5-1						74	38	38	23	14	14	15	21
$\beta 2$							38	37	21	11	12	13	20
KA-1								67	24	14	14	16	20
$\gamma 2$									24	14	14	18	20
$\delta 1$										14	15	17	19
$\epsilon 1$											52	40	18
$\epsilon 2$												39	16
$\epsilon 3$													18

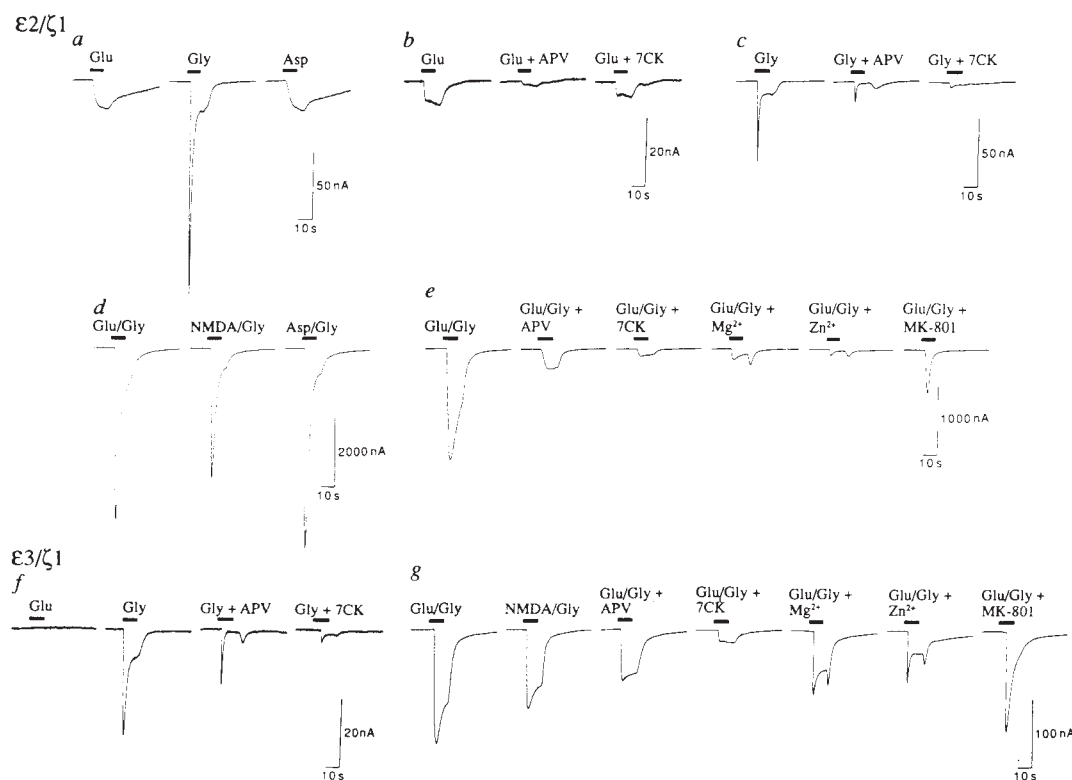
Sequence data are taken from ref. 12 (mouse $\alpha 1$ and $\alpha 2$), ref. 9 (rat GluRC and GluRD), ref. 13 (rat GluR5-1), ref. 15 (mouse $\beta 2$), ref. 16 (rat KA-1), ref. 17 (mouse $\gamma 2$), ref. 20 (mouse $\delta 1$), ref. 21 (mouse $\epsilon 1$) and ref. 19 (mouse $\zeta 1$). Amino-acid sequence identity between GluR channel subunits was calculated using GeneWorks software (IntelliGenetics). Essentially the same values (differences are within 2%) were obtained for the rat counterparts of the mouse $\alpha 1$, $\alpha 2$, $\beta 2$ and $\zeta 1$ subunits, that is, the GluR1, GluRB (GluR2), GluR6 and NMDAR1 subunits.

tively, for oocytes implanted with the $\epsilon 2$ and $\zeta 1$ subunits (Fig. 2d), and 191 ± 67 nA and 89 ± 22 nA ($n = 7$), respectively, for oocytes implanted with the $\epsilon 3$ and $\zeta 1$ subunits (Fig. 2g). The current amplitudes were much larger than those for oocytes implanted with the $\zeta 1$ subunit alone (11 ± 2 nA and 9 ± 2 nA, $n = 12$, respectively). Oocytes injected with the $\epsilon 2$ or $\epsilon 3$ subunit-specific mRNA alone exhibited no detectable response (< 1 nA).

Neither 100 μ M kainate nor 100 μ M AMPA evoked a measurable response of oocytes implanted with the $\epsilon 2$ and $\zeta 1$ subunits and with the $\epsilon 3$ and $\zeta 1$ subunits.

The heteromeric $\epsilon 2/\zeta 1$ channel showed a response to 10 μ M L-glutamate, 10 μ M glycine or 100 μ M L-aspartate applied alone (Fig. 2a), whereas the $\epsilon 3/\zeta 1$ channel exhibited a response only to 10 μ M glycine (Fig. 2f). The responses evoked by 10 μ M

FIG. 2 Current responses of *Xenopus* oocytes implanted with the heteromeric $\epsilon 2/\zeta 1$ and $\epsilon 3/\zeta 1$ NMDA receptor channels. The concentrations of agonists and antagonists are as follows: Glu, 10 μ M L-glutamate; NMDA, 100 μ M; Gly, 10 μ M glycine; Asp, 100 μ M L-aspartate; APV, 100 μ M; 7 CK, 30 μ M; Mg^{2+} , 100 μ M $MgCl_2$; Zn^{2+} , 100 μ M $ZnCl_2$; MK-801, 1 μ M (+)-MK-801. Current responses were measured in normal frog Ringer's solution at -70 mV membrane potential. Inward current is downward. The duration of bath application of agonists and antagonists is indicated by bars without taking into account the dead-space time in the perfusion system (~ 2 s). a, Responses of the $\epsilon 2/\zeta 1$ channel to Glu, Gly and Asp. b, Effects of APV and 7 CK on the response of the $\epsilon 2/\zeta 1$ channel to Glu. c, Effects of APV and 7 CK on the response of the $\epsilon 2/\zeta 1$ channel to Gly. d, Responses of the $\epsilon 2/\zeta 1$ channel to Glu plus Gly, NMDA plus Gly, and Asp plus Gly. e, Effects of APV, 7 CK, Mg^{2+} , Zn^{2+} and MK-801 on the response of the $\epsilon 2/\zeta 1$ channel to Glu plus Gly. f, Responses of the $\epsilon 3/\zeta 1$ channel to Glu and Gly, and effects of APV and 7 CK on the response to Gly. g, Responses of the $\epsilon 3/\zeta 1$ channel to Glu plus Gly, NMDA plus Gly, and effects of APV, 7 CK, Mg^{2+} , Zn^{2+} and MK-801 on the response to Glu plus Gly.



the plasmid pBKSA $\epsilon 3$. The $\epsilon 2$ and $\epsilon 3$ subunit-specific mRNAs were synthesized *in vitro* with T3 RNA polymerase (BRL) in the presence of 0.5 mM cap dinucleotide 7mGpppG using NotI-cleaved pBKSA $\epsilon 2$ and XbaI-cleaved pBKSA $\epsilon 3$ as templates, respectively. *Xenopus laevis* oocytes were injected with the $\epsilon 1$ (~ 22 ng per oocyte, ref. 21), $\epsilon 2$ (~ 19 ng per oocyte), $\epsilon 3$ (~ 16 ng per oocyte) or $\zeta 1$ (~ 13 ng per oocyte, ref. 19) subunit-specific mRNA or in combinations as specified. Whole-cell currents were recorded after 2–3 days incubation with a conventional two-micropipette voltage clamp as described¹².

L-glutamate or 10 μM glycine alone were diminished by 100 μM D-2-amino-5-phosphonovaleate (APV), a specific competitive antagonist of the NMDA receptor²⁹, and 30 μM 7-chlorokynureinate (7 CK), a competitive antagonist for the glycine modulatory site of the NMDA receptor³⁰ (Fig. 2b, c, f); the response of the $\epsilon 2/\zeta 1$ channel to L-glutamate was less sensitive to 7 CK. The responses to 10 μM L-glutamate plus 10 μM glycine were suppressed by these competitive antagonists and noncompetitive antagonists, 100 μM Mg^{2+} , 100 μM Zn^{2+} and 1 μM (+)-MK-801 (Fig. 2e, g). The effects of the channel blockers were apparently weaker for the $\epsilon 3/\zeta 1$ channel than for the $\epsilon 1/\zeta 1$ (ref. 21) and $\epsilon 2/\zeta 1$ channels. The $\epsilon 2/\zeta 1$ and $\epsilon 3/\zeta 1$ channels exhibited clear inward currents in Na^+ - and K^+ -free Ringer's solution containing 20 mM Ca^{2+} (Ca^{2+} -Ringer's solution²⁷), whereas a marginal outward current was observed in control Na^+ - and K^+ -free Ringer's solution, indicating that the heteromeric channels are

permeable to Ca^{2+} . We thus conclude that the $\epsilon 2$ and $\epsilon 3$ proteins are subunits of the NMDA receptor channel.

Dose-response relationships of the heteromeric NMDA receptor channels for L-glutamate and glycine were examined in Ba^{2+} -Ringer's solution to minimize the effect of secondarily activated Ca^{2+} -dependent Cl^- currents³¹. The EC_{50} (effector concentration for half-maximum response) values for L-glutamate were 1.7 μM , 0.8 μM and 0.7 μM for the $\epsilon 1/\zeta 1$, $\epsilon 2/\zeta 1$ and $\epsilon 3/\zeta 1$ channels, respectively, whereas those for glycine were 2.1 μM , 0.3 μM and 0.2 μM , respectively (Fig. 3a, b). Hill coefficient values were 1.2–2.2. The effects of competitive antagonists were investigated at agonist concentrations as high as 10 times the EC_{50} values. The sensitivity to APV is in the order $\epsilon 1/\zeta 1 > \epsilon 2/\zeta 1 > \epsilon 3/\zeta 1$ channels, whereas that to 7 CK is in the order $\epsilon 3/\zeta 1 > \epsilon 2/\zeta 1 > \epsilon 1/\zeta 1$ channels (Fig. 3c). Mg^{2+} at the concentrations of 0.1 mM and 1 mM exerted an inhibitory effect on the ϵ/ζ heteromeric NMDA receptor channels in a voltage-dependent manner (Fig. 3d, e). Clear difference, however, was found among the heteromeric channels. The $\epsilon 3/\zeta 1$ channel was rather resistant to Mg^{2+} block, being active even at -70 mV and -100 mV membrane potentials in the presence of 1 mM Mg^{2+} , whereas both the $\epsilon 1/\zeta 1$ and $\epsilon 2/\zeta 1$ channels were strongly suppressed under these conditions. These results suggest that functionally distinct NMDA receptor channels can be formed depending on the subunit combinations.

Selective expression in the brain

Distributions of the $\epsilon 2$ and $\epsilon 3$ subunit mRNAs in mouse brain were examined by RNA blot hybridization and *in situ* hybridization analyses. RNA of $\sim 20,000$ nucleotides hybridizing with an $\epsilon 2$ subunit-specific probe was found in forebrain RNA but not in cerebellar RNA (Fig. 4a). When exposed longer, a faint signal corresponding to RNA of $\sim 18,000$ nucleotides was detectable in the cerebellar RNA. This may represent an alternatively spliced form of the $\epsilon 2$ subunit mRNA or a cross-hybridizable RNA species. *In situ* hybridization analyses of sagittal and coronal sections revealed intense signals in the pyramidal layer of the hippocampal gyrus, the granular layer of the dentate gyrus, the layers II to VI of the cerebral cortex of the convexity and the olfactory bulb (Fig. 4c, f, g). Among cortical layers, the layer II exhibited stronger signals. Signals were also detected in the caudate-putamen, the thalamic nuclei and the amygdaloid nuclei, but not distinctly in the hypothalamic nuclei. Subcortical white matter and the internal capsule showed no distinct signals. The entire field of the cerebellum and the brainstem including the midbrain, the pons and the medulla oblongata revealed no significant signals.

In contrast, RNA of $\sim 4,600$ and $\sim 12,000$ nucleotides hybridizing with an $\epsilon 3$ subunit-specific probe were found predominantly in cerebellar RNA (Fig. 4b). When exposed longer, a faint signal corresponding to the $\sim 4,600$ nucleotide RNA was found in forebrain RNA. Selective expression of the $\epsilon 3$ subunit mRNA in the cerebellum was confirmed by *in situ* hybridization analyses of sagittal and coronal sections (Fig. 4d, e, h, i). Intense signals were observed in the granule cell layer of the cerebellum, whereas weak signals were detected in olfactory bulb and thalamus. The other regions of the cerebrum and the brainstem revealed no significant signals.

Positive modulation

We have previously shown that the $\zeta 1$ homomeric NMDA receptor channel expressed in *Xenopus* oocytes is transiently potentiated by treatment with TPA¹⁹. Effects on heteromeric channels were examined by measuring responses to 10 μM L-glutamate plus 10 μM glycine in Ba^{2+} -Ringer's solution. Treatment of oocytes with 1 μM TPA for 10 min potentiated the responses of the $\epsilon 2/\zeta 1$ and $\epsilon 1/\zeta 1$ heteromeric channels (Fig. 5), but not the response of the $\epsilon 3/\zeta 1$ channel. The activation of the $\epsilon 2/\zeta 1$ channel was sustained at least for 60 min after washing, whereas the activity of the $\epsilon 1/\zeta 1$ channel reached

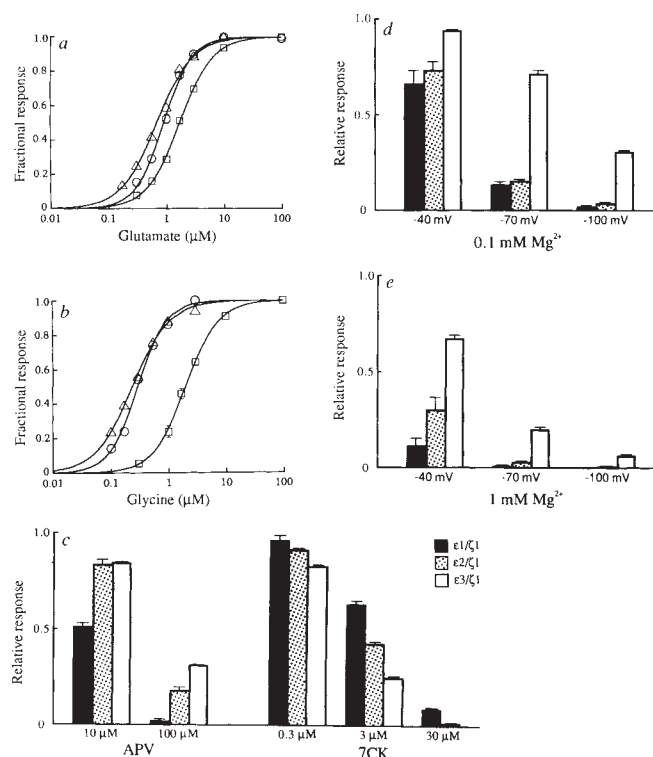


FIG. 3 Functional properties of heteromeric NMDA receptor channels expressed in *Xenopus* oocytes. Current responses were measured at -70 mV membrane potential in Ba^{2+} -Ringer's solution containing 115 mM NaCl, 2.5 mM KCl, 1.8 mM BaCl_2 and 10 mM HEPES-NaOH (pH 7.2), unless otherwise specified. a, b, Dose-response curves for L-glutamate in the presence of 10 μM glycine (a) and for glycine in the presence of 10 μM L-glutamate (b). $\epsilon 1/\zeta 1$ ($\square$); $\epsilon 2/\zeta 1$ ($\circ$); $\epsilon 3/\zeta 1$ ($\triangle$). Each point represents the mean of fractional responses obtained from 3–4 oocytes; s.e.m. are indicated by bars when larger than the symbols. The theoretical curves have been drawn according to the equation $I = I_{\text{max}}/[1 + (\text{EC}_{50}/A)^n]$, where I represents the current response, I_{max} the maximum response, A the concentration of agonist, and n the Hill coefficient. The maximum current responses were 42–70 nA ($\square$), 195–844 nA ($\circ$) and 39–81 nA ($\triangle$) in a, and 26–150 nA ($\square$), 218–472 nA ($\circ$) and 28–49 nA ($\triangle$) in b. c, Effects of APV and 7CK on the response to L-glutamate plus glycine. The concentrations of agonists were 10-fold the respective EC_{50} values. Data are presented as mean $\pm$ s.e.m. of measurements on 3–5 oocytes. Current responses in the absence of antagonists were 20–83 nA, 20–2,390 nA and 20–104 nA for the $\epsilon 1/\zeta 1$, $\epsilon 2/\zeta 1$ and $\epsilon 3/\zeta 1$ channels, respectively. d, e, Effects of 0.1 mM (d) and 1 mM (e) Mg^{2+} on the response to 10 μM L-glutamate plus 10 μM glycine of the $\epsilon 1/\zeta 1$ (filled columns), $\epsilon 2/\zeta 1$ (dotted columns) and $\epsilon 3/\zeta 1$ (open columns) channels at various membrane potentials. Control current responses obtained in the absence of Mg^{2+} at -40 mV, -70 mV and -100 mV membrane potentials were 27–36 nA, 45–69 nA and 56–88 nA for the $\epsilon 1/\zeta 1$ channel, 193–528 nA, 335–778 nA and 618–1,000 nA for the $\epsilon 2/\zeta 1$ channel, and 10–13 nA, 22–23 nA and 34–43 nA for the $\epsilon 3/\zeta 1$ channel, respectively.

maximum level after 20 min and gradually decreased to the initial level after 60 min. No significant enhancement of the channel activity was observed by mock treatment or TPA treatment in the presence of 5 μ M staurosporine, a protein kinase inhibitor³². These results suggest that the $\epsilon 1/\zeta 1$ and $\epsilon 2/\zeta 1$ heteromeric NMDA receptor channels can be positively modulated directly or indirectly by protein kinases and that the duration of modulation varies with the ϵ subunit combined.

Discussion

We have previously demonstrated that NMDA receptor channels with high activity are formed by coexpression of the dis-

tantly related $\epsilon 1$ and $\zeta 1$ subunits²¹. Here we have identified two additional members of the ϵ subfamily of the GluR channel and have shown that both the $\epsilon 2$ and $\epsilon 3$ subunits together with the $\zeta 1$ subunit yield GluR channels with characteristics of the NMDA receptor channel such as requirement for glycine and inhibition or blocking by APV, 7CK, Mg^{2+} , Zn^{2+} and (+)-MK-801. The $\epsilon 2$ and $\epsilon 3$ subunits are highly homologous in primary structure to the $\epsilon 1$ subunit but are clearly distinct in distribution, functional properties and regulation. In contrast to the wide distribution of the $\epsilon 1$ and $\zeta 1$ subunit mRNAs in various neural cells²¹, the $\epsilon 2$ subunit mRNA is expressed selectively in the forebrain, and the $\epsilon 3$ subunit mRNA is found predominantly

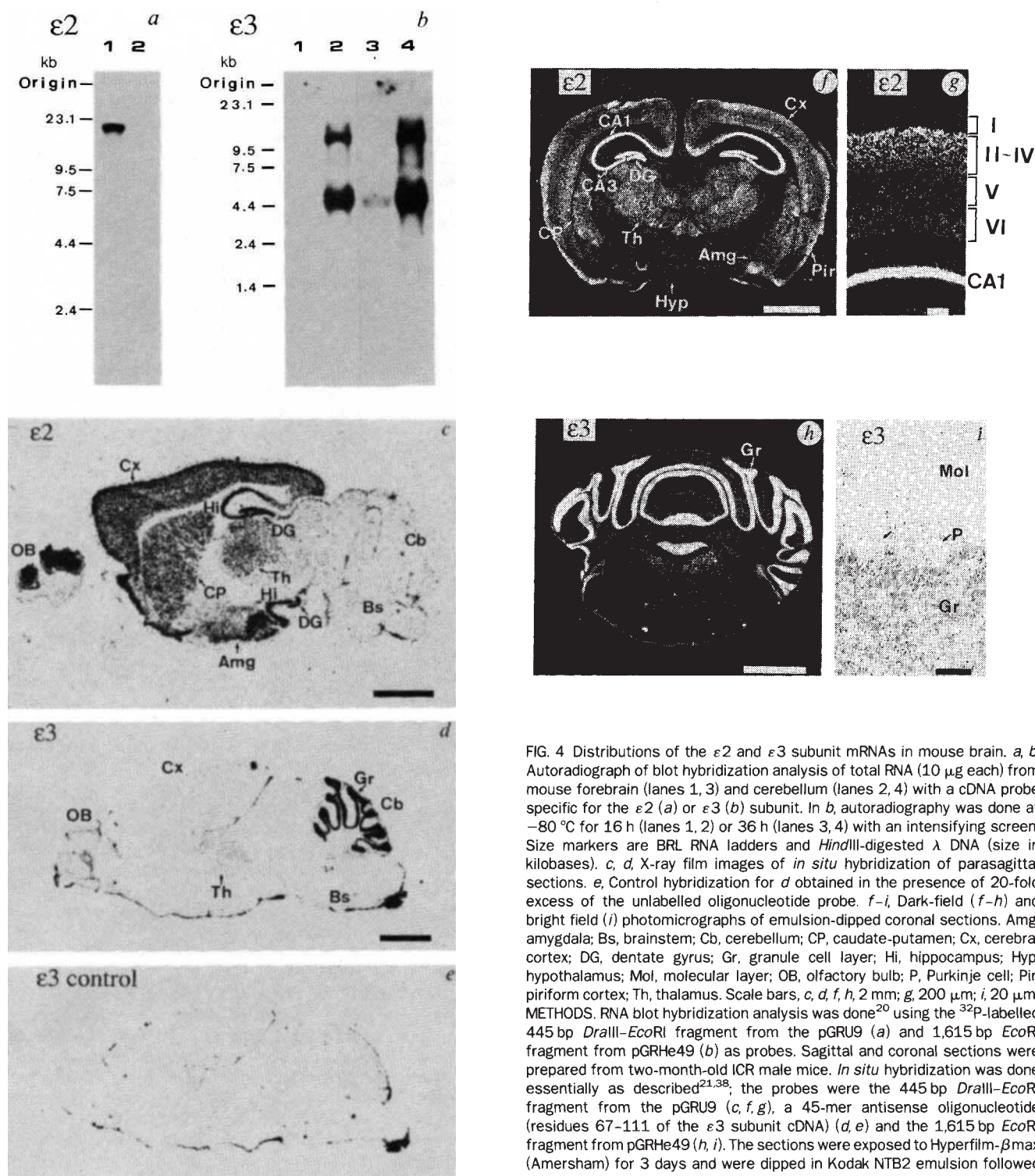


FIG. 4 Distributions of the $\epsilon 2$ and $\epsilon 3$ subunit mRNAs in mouse brain. *a, b*, Autoradiograph of blot hybridization analysis of total RNA (10 μ g each) from mouse forebrain (lanes 1, 3) and cerebellum (lanes 2, 4) with a cDNA probe specific for the $\epsilon 2$ (*a*) or $\epsilon 3$ (*b*) subunit. In *b*, autoradiography was done at -80°C for 16 h (lanes 1, 2) or 36 h (lanes 3, 4) with an intensifying screen. Size markers are BRL RNA ladders and *Hind*III-digested λ DNA (size in kilobases). *c, d*, X-ray film images of *in situ* hybridization of parasagittal sections. *e*, Control hybridization for *d* obtained in the presence of 20-fold excess of the unlabelled oligonucleotide probe. *f-i*, Dark-field (*f-h*) and bright field (*i*) photomicrographs of emulsion-dipped coronal sections. Amg, amygdala; Bs, brainstem; Cb, cerebellum; CP, caudate-putamen; Cx, cerebral cortex; DG, dentate gyrus; Gr, granule cell layer; Hi, hippocampus; Hyp, hypothalamus; Mol, molecular layer; OB, olfactory bulb; P, Purkinje cell; Pir, piriform cortex; Th, thalamus. Scale bars, *c, d, f, h*, 2 mm; *g*, 200 μ m; *i*, 20 μ m. METHODS. RNA blot hybridization analysis was done²⁰ using the ³²P-labelled 445 bp *Dra*III-*Eco*RI fragment from the pGRU9 (*a*) and 1,615 bp *Eco*RI fragment from pGRHe49 (*b*) as probes. Sagittal and coronal sections were prepared from two-month-old ICR male mice. *In situ* hybridization was done essentially as described^{21,38}, the probes were the 445 bp *Dra*III-*Eco*RI fragment from the pGRU9 (*c, f, g*), a 45-mer antisense oligonucleotide (residues 67-111 of the $\epsilon 3$ subunit cDNA) (*d, e*) and the 1,615 bp *Eco*RI fragment from pGRHe49 (*h, i*). The sections were exposed to Hyperfilm- β max (Amersham) for 3 days and were dipped in Kodak NTB2 emulsion followed by exposure for 2-4 weeks.

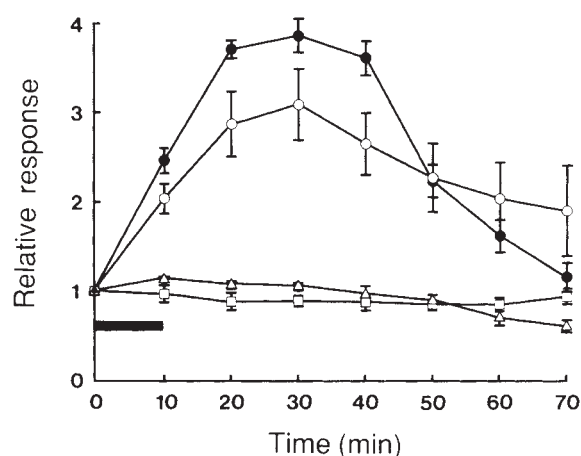


FIG. 5 Effect of TPA treatment. Whole-cell currents activated by bath-application of 10 μ M L-glutamate plus 10 μ M glycine before and after treatment with 1 μ M TPA for 10 min were recorded at -70 mV membrane potential in Ba^{2+} -Ringer's solution from oocytes injected with the $\epsilon 2$ subunit-specific and $\zeta 1$ subunit-specific mRNAs (○) or with the $\epsilon 1$ subunit-specific and $\zeta 1$ subunit-specific mRNAs (●). For control, injected oocytes were treated with 0.005% dimethyl sulphoxide alone (mock treatment, □) or with 1 μ M TPA and 5 μ M staurosporine (△). The duration of treatment is indicated by a thick bar. Each point represents the mean of relative responses obtained from 3 to 5 oocytes; s.e.m. are indicated by bars when larger than the symbols.

in the cerebellum. The $\epsilon 2/\zeta 1$ and $\epsilon 3/\zeta 1$ channels exhibit higher apparent affinities for L-glutamate and glycine than the $\epsilon 1/\zeta 1$ channel. Furthermore, the $\epsilon 1/\zeta 1$ channel is most sensitive to APV, whereas the $\epsilon 3/\zeta 1$ channel is most sensitive to 7CK. A remarkable feature of the $\epsilon 3/\zeta 1$ channel is its resistance to Mg^{2+} block in contrast to the higher sensitivity of the $\epsilon 1/\zeta 1$ and $\epsilon 2/\zeta 1$ channels. These findings suggest that the functional properties of the NMDA receptor channel are critically determined by the constituting ϵ subunit. Because the activity-dependent change in synaptic efficacy mediated by the NMDA receptor channel is found in forebrain regions such as the hippocampus and the visual cortex^{1,2,4}, it is likely that the $\epsilon 2$ subunit exclusively expressed in the forebrain plays an important role in the synaptic plasticity. In contrast, the distribution of the $\epsilon 3$ subunit mRNA and the resistance of the $\epsilon 3/\zeta 1$ channel to Mg^{2+} block suggest that the $\epsilon 3$ subunit is a key component of cerebellar NMDA receptor channels mediating mossy fibre-granule cell synaptic transmission^{23,33}. Radioligand binding studies have differentiated three populations of NMDA receptors with different pharmacological characteristics and regional distribution in the brain^{24,25}. Furthermore, patch-clamp studies comparing different types of cerebellar neurons suggest functional diversity

of the NMDA receptor channel, although no specific features were noted for cerebellar channels³³⁻³⁵. The molecular diversity of the ϵ subunit family probably underlies such pharmacological and electrophysiological heterogeneity of the NMDA receptor channel.

We have also shown that treatment with TPA enhances the channel activity of the $\epsilon 1/\zeta 1$ and $\epsilon 2/\zeta 1$ channels, but not the $\epsilon 3/\zeta 1$ channel. Once potentiated, the activation of the $\epsilon 2/\zeta 1$ channel was sustained at least for one hour, whereas the activity of the $\epsilon 1/\zeta 1$ channel returned to the initial level. Failure of potentiation in the presence of staurosporine suggests the involvement of protein kinases in the TPA effect. These observations raise an intriguing possibility that NMDA receptor channels can be positively modulated by protein kinases. Thus the level of NMDA receptor channel activity may be regulated by various stimuli leading to activation of protein kinases, for example through G protein-coupled receptors. Such modulation may play a role in determining the threshold of the induction of long-term potentiation, because Ca^{2+} entry through the NMDA receptor channel triggers the persistent change in the efficacy of synaptic transmission^{1,2}. □

Received 30 March; accepted 28 May 1992.

1. Mayer, M. L. & Westbrook, G. L. *Prog. Neurobiol.* **28**, 197-276 (1987).
2. Collingridge, G. L. & Bliss, T. V. P. *Trends Neurosci.* **10**, 288-293 (1987).
3. Ito, M. *A. Rev. Neurosci.* **12**, 85-102 (1989).
4. McDonald, J. W. & Johnston, M. V. *Brain Res. Rev.* **15**, 41-70 (1990).
5. Choi, D. W. *Neuron* **1**, 623-634 (1988).
6. Olney, J. W. *A. Rev. Pharmac. Tox.* **30**, 47-71 (1990).
7. Monaghan, D. T., Bridges, R. J. & Cotman, C. W. *A. Rev. Pharmac. Tox.* **29**, 365-402 (1989).
8. Hollmann, M., O'Shea-Greenfield, A., Rogers, S. W. & Heinemann, S. *Nature* **342**, 643-648 (1989).
9. Keinänen, K. *et al. Science* **249**, 556-560 (1990).
10. Boulter, J. *et al. Science* **249**, 1033-1037 (1990).
11. Nakanishi, N., Schneider, N. A. & Axel, R. *Neuron* **5**, 569-581 (1990).
12. Sakimura, K. *et al. FEBS Lett.* **272**, 73-80 (1990).
13. Bettler, B. *et al. Neuron* **5**, 583-595 (1990).
14. Egebjerg, J., Bettler, B., Hermans-Borgmeyer, I. & Heinemann, S. *Nature* **351**, 745-748 (1991).
15. Morita, T. *et al. Molec. Brain Res.* **14**, 143-146 (1992).
16. Werner, P., Voigt, M., Keinänen, K., Wisden, W. & Seeburg, P. H. *Nature* **351**, 742-744 (1991).
17. Sakimura, K., Morita, T., Kushiya, E. & Mishina, M. *Neuron* **8**, 267-274 (1992).
18. Moriyoshi, K. *et al. Nature* **354**, 31-37 (1991).
19. Yamazaki, M., Mori, H., Araki, K., Mori, K. J. & Mishina, M. *FEBS Lett.* **300**, 39-45 (1992).
20. Yamazaki, M., Araki, K., Shibata, A. & Mishina, M. *Biochem. biophys. Res. Commun.* **183**, 886-892 (1992).
21. Meguro, H. *et al. Nature* **357**, 70-74 (1992).
22. Puckett, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 7557-7561 (1991).

23. Garthwaite, J. & Brodbelt, A. R. *Neuroscience* **29**, 401-412 (1989).
24. Monaghan, D. T. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 9836-9840 (1988).
25. Monaghan, D. T. & Anderson, K. J. in *Excitatory Amino Acids and Synaptic Transmission* (eds Wheal, H. & Thomson, A.) 33-54 (Academic, San Diego, 1991).
26. Hume, R. I., Dingledine, R. & Heinemann, S. *Science* **253**, 1028-1031 (1991).
27. Mishina, M. *et al. Biochem. biophys. Res. Commun.* **180**, 813-821 (1991).
28. Burnashev, N., Monyer, H., Seeburg, P. H. & Sakmann, B. *Neuron* **8**, 189-198 (1992).
29. Davies, J., Francis, A. A., Jones, A. W. & Watkins, J. C. *Neurosci. Lett.* **21**, 77-81 (1981).
30. Kemp, J. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6547-6550 (1988).
31. Leonard, J. P. & Kelso, S. R. *Neuron* **4**, 53-60 (1990).
32. Tamaoki, T. *et al. Biochem. biophys. Res. Commun.* **135**, 397-402 (1986).
33. Silver, R. A., Traynelis, S. F. & Cull-Candy, S. G. *Nature* **355**, 163-166 (1992).
34. Howe, J. R., Cull-Candy, S. G. & Colquhoun, D. *J. Physiol.* **432**, 143-202 (1991).
35. Traynelis, S. F. & Cull-Candy, S. G. *J. Physiol.* **433**, 727-763 (1991).
36. von Heijne, G. *Nucleic Acids Res.* **14**, 4683-4690 (1986).
37. Pearson, R. B., Woodgett, J. R., Cohen, P. & Kemp, B. E. *J. Biol. Chem.* **260**, 14471-14476 (1985).
38. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
39. Usui, H. *et al. Molec. chem. Neuropath.* **15**, 207-216 (1991).

ACKNOWLEDGEMENTS. We thank M. Watanabe and T. Ichikawa for sections, M. Nakazawa for advice, M. Yamazaki for help in cDNA cloning and M. Motokoshi for help in the preparation of the manuscript. This investigation was supported in part by grants from the Ministry of Education, Science and Culture of Japan, the Institute of Physical and Chemical Research, the Ministry of Health and Welfare of Japan, and the Yujin Memorial Grant.

Is a pyrene-like molecular ion the cause of the 4,430-Å diffuse interstellar absorption band?

F. Salama & L. J. Allamandola

NASA-Ames Research Center, Space Science Division,
MS 245-6 Moffett Field, California 94035, USA

THE diffuse interstellar bands (DIBs), ubiquitous absorption features in astronomical spectra, have been known since early this century¹ and now number more than a hundred. Ranging from 4,400 Å to the near infrared², they differ markedly in depth, width and shape, making the concept of a single carrier unlikely. Whether they are due to gas or grains is not settled, but recent results³ suggest that the DIB carriers are quite separate from the grains that cause visual extinction. Among molecular candidates the polycyclic aromatic hydrocarbons (PAHs) have been proposed as the possible carriers of some of the DIBs⁴⁻⁷, and we present here laboratory measurements of the optical spectrum of the pyrene cation $C_{16}H_{10}^+$ in neon and argon matrices. The strongest absorption feature falls at $4,435 \pm 5$ Å in the argon matrix and $4,395 \pm 5$ Å in the neon matrix, both close to the strong 4,430-Å DIB. If this or a related pyrene-like species is responsible for this particular band, it must account for 0.2% of all cosmic carbon. The ion also shows an intense but puzzling broad continuum, extending from the ultraviolet to the visible, similar to what is seen in the naphthalene cation⁸ and perhaps therefore a common feature of all PAH cations. This may provide an explanation of how PAHs convert a large fraction of interstellar radiation from ultraviolet and visible wavelengths down to the infrared.

PAHs are thought to be ubiquitous and abundant in the interstellar medium^{7,9} and are stable against ultraviolet photodissociation, so they are attractive candidates for the DIB carriers. Moreover, a large fraction of the PAHs are expected to be ionized in the interstellar medium⁴⁻⁷, and thus absorb lower-energy photons (mainly in the visible and near-infrared regions of the spectrum) than the neutral precursors¹⁰.

The only available data on PAHs come from absorption spectra of neutral and ionized PAHs suspended in perturbing media (solid phase¹⁰⁻¹² or solution¹³) or from gas-phase photoelectron spectra which do not provide information on all the possible optical transitions¹⁴. We have initiated a program to measure the spectroscopic properties of neutral and ionized PAHs in the ultraviolet, visible and near-infrared range (1,800–9,000 Å) under conditions relevant to astrophysical environments (isolated species in the least-perturbing solid media possible, inert gas matrices). We have already reported the detailed spectroscopic analysis of the smallest PAH, naphthalene ($C_{10}H_8$) and its positive ion ($C_{10}H_8^+$)¹⁵, and discussed its relevance to astrophysical questions⁸. Here we consider the possible contribution of the ion $C_{16}H_{10}^+$ to the 4,430-Å DIB and we describe the broad, structureless continuum, extending from 2,000 to 5,100 Å, which seems to be associated with it.

The experimental apparatus, described in detail in ref. 15, consists of a computer-controlled ultraviolet-visible-near-infrared spectrometer system, coupled to a helium-cooled cryogenic cell. Matrix isolation spectroscopy, which consists of isolating the species of interest ($C_{16}H_{10}$, $C_{16}H_{10}^+$) in a chemically inert lattice, is particularly adapted to astrophysical applications as it can be used to trap ions with minimum perturbation to their electronic spectrum^{15,16}. Argon is more convenient than neon because of its higher sublimation temperature, and is used to explore the experimental parameters important to optimize ion yield. Neon, however, is the least-perturbing solid medium known, generally producing shifts in vibronic positions of a few tenths of a per cent with respect to the gas phase¹⁶. We therefore focus on the neon matrix results, as they should be more relevant to astrophysics.

In the wavelength range 1,800–9,000 Å, neutral pyrene trapped in a neon matrix shows only three distinct band systems. The strongest absorptions peak in the ranges 2,320–2,350 Å, 2,655–2,675 Å and 3,247–3,273 Å, respectively, as a function of concentration (F.S. and L.J.A., manuscript in preparation). All these bands have similar intensities (oscillator strength $f \approx 0.5$) and thus can be searched for on the extinction curve. The shortest-wavelength neutral pyrene band may, of course, blend with the broad 2,175-Å extinction hump. The cross-sections of these bands (1.5×10^{-16} to 0.9×10^{-16} cm² molecule⁻¹)¹³ require a neutral pyrene abundance of 1.5×10^{-7} with respect to hydrogen to produce substructure on the extinction curve at the 10% level. This is high compared with the overall PAH abundance, thought to range from 2.0×10^{-7} to 5.4×10^{-7} (refs 7, 9). Furthermore, the fraction of pyrene in the neutral form should be small in the general interstellar medium and depends on the ionization balance along the specific line of sight. Figure 1a shows the visible part of the spectrum (3,800–7,800 Å). From the astrophysical perspective, this spectrum indicates that neutral pyrene does not absorb in the visible and therefore cannot contribute to the DIBs.

Vacuum ultraviolet irradiation of neutral $C_{16}H_{10}$ isolated in a rare-gas matrix with 10.2-eV photons produces the pyrene cation ($C_{16}H_{10}^+$) by direct one-photon ionization. The entire ultraviolet-visible-near-infrared spectrum of $C_{16}H_{10}^+$ is largely dominated by a strong single absorption which falls at $4,395 \pm 5$ Å in neon and $4,435 \pm 5$ Å in argon matrices (Fig. 1b and c). A continuum, extending from 2,000 to 5,100 Å (not all shown here), is also produced. The new spectral features that arise on vacuum ultraviolet irradiation of neutral pyrene are attributed to the pyrene cation because their positions and relative intensities are similar to those previously assigned to this ion and measured with different cation formation techniques¹⁰⁻¹⁴. We will limit our discussion here to the strongest absorption feature of $C_{16}H_{10}^+$, which may be connected with the strong 4,430-Å DIB. A detailed spectroscopic study of the entire ultraviolet-visible spectrum of $C_{16}H_{10}^+$ is in progress (F.S. and L.J.A., manuscript in preparation). The effect of matrix material on band positions is shown by the redshift of 40 Å measured when going from neon to argon (Fig. 2). Thus, the good agreement observed between the 4,435-Å absorption of $C_{16}H_{10}^+$ in argon and the 4,430-Å DIB, although suggestive, is probably coincidental (induced by the matrix effect). Nonetheless, the results raise the interesting question of whether the proximity of the 4,395 Å absorption of $C_{16}H_{10}^+$ to the 4,430-Å DIB is a coincidence, or an indication that the carrier of the 4,430-Å DIB may be the pyrene cation ($C_{16}H_{10}^+$) or a closely related, pyrene-

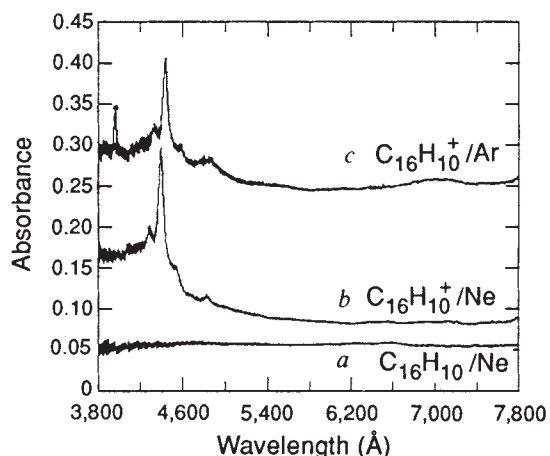


FIG. 1 Visible absorption spectra of (a) pyrene ($C_{16}H_{10}$) isolated in a neon matrix at 4.2 K; (b) the pyrene cation ($C_{16}H_{10}^+$) in a neon matrix at 4.2 K and (c) the pyrene cation ($C_{16}H_{10}^+$) in an argon matrix at 4.2 K. The dot (●) indicates a band arising from an impurity in the sample.

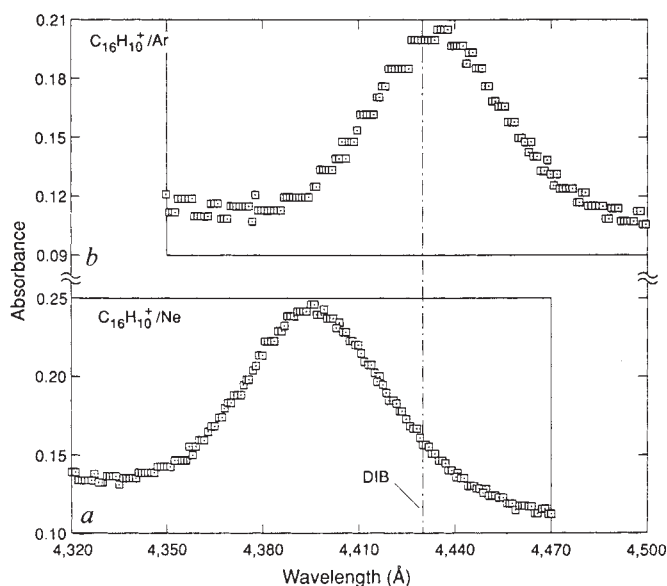


FIG. 2 Profile and central wavelength of the strongest visible absorption band of the pyrene cation ($C_{16}H_{10}^+$) when isolated, at 4.2 K, in (a) a neon matrix and (b) an argon matrix.

like species. Side-groups (such as CH_2 and CH_3) are known to induce a redshift of the vibronic transitions of other PAH cations¹⁷. This 40-Å shift also illustrates the important, well known influence of the solid medium on band position. If any DIBs originate in a solid material, any significant compositional variation from one line-of-sight to another should cause detectable wavelength shifts. No such shifts have been found, despite considerable observational efforts. This supports the idea that those DIBs which do not vary in position from one line-of-sight to another arise from free gas-phase molecules.

The 4,430-Å interstellar absorption, when observed, is the strongest and the broadest (full width half maximum ~ 20 Å) of the known DIBs¹. It is also highly variable in depth and is weaker towards objects with strong far ultraviolet extinction¹⁸. We have shown⁸ that a small number of DIBs (six in all) coincide with bands of the naphthalene cation ($C_{10}H_8^+$), and we argued that a mixture of different PAH ions may well explain some (or most) of the DIBs. This suggestion seems to be supported by the present results on the pyrene cation. The answer to this problem can only be obtained through a detailed and systematic comparison of the DIBs with appropriate laboratory data on a series of PAHs. The $C_{16}H_{10}^+$ oscillator strength of 0.126 in argon implies that 0.2% of the cosmic carbon must be in the form of $C_{16}H_{10}^+$ if the pyrene cation is responsible for the 4,430-Å DIB. This is comparable to the 0.3% abundance derived for the naphthalene cation⁸.

Another important result of these experiments is the production of a broad and strong continuum extending from the ultraviolet to the visible (2,000–3,600 Å with a weak tail extending to $\sim 5,100$ Å), which seems to be correlated with the discrete features associated with the pyrene cation (see Fig. 1). A similar continuum has been reported for the naphthalene cation^{8,15}. Although the nature of the transition responsible for this continuum in $C_{16}H_{10}^+$ (and $C_{10}H_8^+$) is not known, other possible origins including solvated electrons and PAH anions have been eliminated on the basis of previous experiments^{8,15,19}. Of course, overlapping broad continua will not produce narrow structure on the extinction curve, but they could contribute to the 'very broad structure'^{20,21}. If a broad ultraviolet-visible continuum is indeed an intrinsic spectroscopic property of free, gas-phase PAH cations and not a matrix effect, it could provide the channel by which interstellar ultraviolet-visible radiation is converted to the discrete infrared emission bands in many objects, as most

of PAHs in the infrared-emitting regions are ionized⁷. The ability of the infrared-emitting material to absorb in the visible is required to account for the infrared emission intensities from several objects^{22–24} and constitutes a key test for the validity of the PAH model^{7,9}. □

Received 10 March; accepted 1 May 1992.

- Herbig, G. H. *Astrophys. J.* **196**, 129–160 (1975).
- Herbig, G. H. *Astrophys. J.* **382**, 193–203 (1991).
- Snow, T. P. & Seab, C. G. *Astrophys. J.* **382**, 189–192 (1991).
- Van der Zwet, G. P. & Allamandola, L. J. *Astr. Astrophys.* **146**, 76–80 (1985).
- Leger, A. & d'Hendecourt, L. B. *Astr. Astrophys.* **146**, 81–85 (1985).
- Crawford, M. K., Tielens, A. G. G. M. & Allamandola, L. J. *Astrophys. J.* **293**, L45–L48 (1985).
- Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. *Astrophys. J.* **290**, L25–L28 (1985).
- Salama, F. & Allamandola, L. J. *Astrophys. J.* (in the press).
- Léger, A. & Puget, J. L. *Astr. Astrophys.* **137**, L5–L8 (1984).
- Shida, T. & Iwata, S. *J. Am. chem. Soc.* **95**, 3473–3483 (1973).
- Pankasem, S. & Thomas, J. K. *J. phys. Chem.* **95**, 7385–7393 (1991).
- Kira, A., Imamura, M. & Shida, T. *J. phys. Chem.* **80**, 1445–1448 (1976).
- Karcher, W., Fordham, R. J., Dubois, J. J., Glaude, P. G. J. M. & Lighthart, J. A. M. in *Spectral Atlas of Polycyclic Aromatic Compounds* (ed. Commission of European Communities) 92–95 (Reidel, Dordrecht, 1983).
- Clair, E. & Schmidt, W. *Tetrahedron* **32**, 2563–2566 (1976).
- Salama, F. & Allamandola, L. J. *J. chem. Phys.* **94**, 6964–6977 (1991).
- Bondybey, V. E. & Miller, T. A. in *Molecular Ions: Spectroscopy, Structure and Chemistry* (eds Miller, T. A. & Bondybey, V. E.) 125–173 (North-Holland, New York, 1983).
- Andrews, L., Kelsall, B. J. & Blankenship, T. A. *J. phys. Chem.* **86**, 2916–2926 (1982).
- Krelowski, J. in *Interstellar Dust* (eds Allamandola, L. J. & Tielens, A. G. G. M.) 67–86 (Kluwer, Dordrecht, 1989).
- Salama, F. & Allamandola, L. J. *J. chem. Phys.* **95**, 6190–6191 (1991).
- Hayes, D. S., Mavko, G. E., Radick, R. R., Rex, K. H. & Greenberg, J. M. in *Interstellar Dust and Related Topics* (eds Greenberg, J. M. & Van de Hulst, H. C.) 83–90 (Reidel, Dordrecht, 1973).
- Van Breda, I. G. & Whittet, D. C. B. *Mon. Not. R. astr. Soc.* **195**, 79–88 (1981).
- Sellgren, K. *Astrophys. J.* **277**, 623–633 (1984).
- Sellgren, K., Luan, L. & Wener, M. W. *Astrophys. J.* **359**, 384–391 (1990).
- Aitken, D. K. & Roche, P. F. *Mon. Not. R. astr. Soc.* **202**, 1233–1244 (1983).

ACKNOWLEDGEMENTS. We thank Bob Walker for technical support and Ted Snow for discussions, especially regarding the matrix shift and a grain DIB origin. This work was supported by NASA.

Possible cometary origin of heavy noble gases in the atmospheres of Venus, Earth and Mars

Tobias Owen*, Akiva Bar-Nun† & Idit Kleinfeld†

* University of Hawaii, Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

† Department of Geophysics and Planetary Sciences, Tel Aviv University, Ramat Aviv, 69978 Tel Aviv, Israel

MODELS that trace the origin of noble gases in the atmospheres of the terrestrial planets (Venus, Earth and Mars) to the 'planetary component' in chondritic meteorites confront several problems. The 'missing' xenon in the atmospheres of Mars and Earth (Fig. 1) is one of the most obvious; this gas is not hidden or trapped in surface materials¹. On Venus, the absolute abundances of neon and argon per gram of rock are higher even than those in carbonaceous chondrites, whereas the relative abundances of argon and krypton are closer to solar than to chondritic values (there is only an upper limit on xenon)² (Fig. 1). Pepin³ has developed a model that emphasizes hydrodynamic escape of early, massive hydrogen atmospheres to explain the abundances and isotope ratios of noble gases on all three planets. We have previously suggested that the unusual abundances of heavy noble gases on Venus might be explained by the impact of a low-temperature comet^{4–6}. Further consideration of the probable history of the martian atmosphere⁷, the noble-gas data from the (Mars-derived) SNC meteorites^{8–10} and laboratory experiments on the trapping of noble gases in ice^{6,11,12} lead us to propose here that the noble gases in the atmospheres of all of the terrestrial planets are dominated by a mixture of an internal component and a contribution from impacting icy planetesimals (comets). If true, this hypothesis illustrates the importance of impacts in determining the volatile inventories of these planets.

In our model, hydrodynamic escape would play a role in atmospheric evolution, particularly affecting the fractionation of neon, but might be less important than Pepin's model³ suggests, as the comets would deliver an already fractionated mixture of gas. The giant impacts seen by Wetherill¹³ as an essential part of the planet-building process are an integral feature of our proposal because they erode the atmospheres without fractionation^{14,15}.

We begin by replotting the data for the heavy noble gases in Fig. 1 on a three-isotope plot (Fig. 2a). In this kind of plot, a family of objects containing gases from two reservoirs with differing isotope abundances will lie along the line connecting the isotope ratios in these reservoirs. The isolation of Earth, Mars and Venus from the chondritic meteorites is clearly demonstrated, as is the difference between Venus and the other two planets. The positions of Earth and Mars take on added significance if we examine a similar plot⁹ that includes noble gas isotope ratios from the SNC meteorites (Fig. 2b). As Ott and Begemann⁹ pointed out, the relative abundances of argon, krypton and xenon in different SNC meteorites lie along a line that neatly includes the atmospheric values of these gases on Mars and Earth. It is now generally assumed that the SNC meteorites come from Mars^{3,8,16-18}. The line that these meteorites define in Fig. 2b may therefore be regarded as a mixing line between martian mantle rock (represented by Chassigny) and martian atmosphere (gas trapped in the glassy inclusions of EETA 79001). Although Ott and Begemann⁹ and Ott¹⁰ raised some objections to this interpretation, it seems possible to satisfy their main concerns^{7,19-21}.

How then to explain the position of Earth's atmosphere still farther up the line? The most straightforward interpretation is to assume that the atmospheres of both Mars and Earth contain a contribution from some external, volatile-rich source, and that the Earth now shows a higher proportion of this contribution than does Mars. But what is this source? Neither the chondritic meteorites nor the solar wind will do. A number of authors²²⁻²⁵ have suggested that comets brought volatiles to the early Earth, but definitive evidence has been lacking. Recent experiments on the trapping of gases by amorphous ice^{6,11,12} now seem to

provide that evidence, and implicate icy planetesimals as the principal external source of inner planet volatiles.

This is illustrated in Fig. 2b, where triangles correspond to the noble gas abundance ratios trapped by ice forming at various temperatures. Fractionation occurs for temperatures above 30 K because the differing sizes and polarizabilities of the atoms lead to competition for the available sites in the ice. Note that ice deposited at a temperature of ~50 K traps the heavy noble gases in proportions that lie on an extrapolation of the SNC-Mars-Earth mixing line. This temperature is certainly representative of the temperatures at which comets are presumed to have formed in the Uranus-Neptune region of the solar nebula²⁶. Theoretical work on the recondensation of icy grains²⁷ has re-affirmed the similarity of the laboratory experiments to conditions and processes at this location in the early solar nebula. Furthermore, 50 K is essentially identical to the formation temperature for Halley's comet deduced²⁸ from its parent CO abundance. It therefore seems that comets are a credible source for the gas mixture that forms the true upper end point of the mixing line.

This idea can be tested further by examining gases from the Earth's interior. If the model is correct, then we could expect noble gas abundances measured in mantle-derived rocks on Earth to fall along the same Mars-Earth mixing line, below the point corresponding to the Earth's atmosphere. Gases trapped in such rocks should consist of a mixture of the externally supplied reservoir (from icy planetesimals) and the internal reservoir (gases trapped in the rocks that formed the planet). We expect the internal reservoir to become increasingly dominant as we examine gases from rocks that have been less exposed to the atmosphere, which is itself a mixture of the two reservoirs. Figure 2b shows that this is indeed the case for gases measured in the Loihi seamount²⁹, an average of mid-ocean ridge basalts (MORB)³⁰, and the Reykjanes ridge³¹.

Compared with Mars, the terrestrial atmospheric reservoir has been less depleted by impact after emplacement, as demonstrated by the larger abundances of noble gases per gram of planet (Fig. 1), the lower abundances of nucleogenic gases in the atmosphere compared with the rocks and the location of the Ar/Xe-Kr/Xe point on the mixing line (Fig. 2b). Because our model assumes that the planetary heavy noble gases consist of a mixture of a rocky (internal) reservoir and an ice-delivered (external) reservoir, removing the atmosphere by impact and allowing it to reconstitute itself through outgassing from the planet's interior will have the effect of moving the atmosphere point down the mixing line from its original position. If these impacts occurred sufficiently early in the planet's outgassing history, they will have depleted the non-nucleogenic gases relative to those produced by radioactive decay in crustal rocks and later released⁷. This effect is augmented in the case of ¹²⁹Xe by the differential solubility of iodine and xenon in water, as Musselwhite *et al.*²¹ have pointed out.

The Venus noble gas data define a diagonal domain that continues off the edge of Fig. 2b, owing to present uncertainties on argon and krypton values and the upper limit on xenon. The comet-impact model just described would imply that Venus received a mixture of heavy noble gases similar to the one that contributes to the present atmospheres of Mars and Earth. Evidently this cometary contribution was overwhelmed by something else, either an original component of nearly solar composition trapped in the rocks of the planet itself³² or an external contribution brought by an object distinctly different from those already identified in Fig. 2b.

There are insufficient data at present to decide between these two possibilities. Wetherill¹³ has argued that the prevalence of giant impacts among the terrestrial planets will tend to prevent any one of them from retaining a unique chemical composition. (The non-terrestrial value of the oxygen isotopes in SNC meteorites³³ weakens this argument.) Yet on Venus, not only is the heavy noble gas abundance pattern different from that on

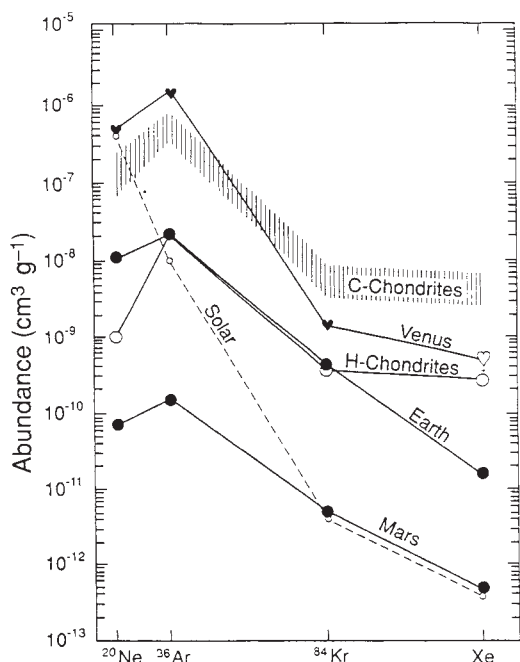


FIG. 1. Abundances of noble gases in planetary atmospheres and chondritic meteorites given as cubic centimetres per gram of rock (after E. Anders, personal communication).

Earth and Mars, but the abundances of both argon and neon are remarkably high. This observation seems to argue for an external contribution after the principal impact phase was over, or for fewer atmosphere-eroding impacts than the Earth experienced. A relatively late impact by a large comet (or group of comets) formed at a temperature near or below 30 K would solve this problem by bringing in a nearly solar mixture of the noble gases (Fig. 2b) after the expected early giant impacts of rocky planetesimals¹³. This would simultaneously explain the unusually large abundance of argon and the 'non-planetary' abundance pattern.

The sizes of these hypothetical comets are not extreme. Assuming for illustrative purposes that only one comet was responsible in each case, we require a diameter of 120 km for

the comet that struck Venus and only 80 km for the object that struck the Earth. These values may be compared with the ~200 km diameter of Chiron, which is now known to be a cometary object³⁴⁻³⁶.

There remains the problem of neon, as this gas is not trapped in amorphous ice except at temperatures below 20 K (refs 6, 11, 12). We propose that the neon in all three atmospheres is a relic of that originally trapped in the rocks that formed the planets. This was originally suggested³⁷ to explain the presence of neon with super-atmospheric $^{20}\text{Ne}/^{22}\text{Ne}$ in Kilauea fumarole gas, and continues to be a viable hypothesis for the similar neon found in Loihi and Kilauea basalts³⁸. This association of high $^{20}\text{Ne}/^{22}\text{Ne}$ with low $^{36}\text{Ar}/^{132}\text{Xe}$ in mantle-derived rocks strengthens our hypothesis. In both cases, we seem to be seeing evidence for the presence of gases that were originally implanted in the rocks that make up the planet. The implantation processes for neon and the heavier gases may have been different, however, so we should not expect a tight correlation. Global loss of neon from the atmosphere by impact erosion^{14,15} and hydrodynamic escape³ could lead to the observed depletion and isotope fractionations for Venus, Earth and Mars. The impact erosion that reduced the heavy noble gases to their present abundances (per gram of planet, see Fig. 1) would reduce the neon as well, without changing $^{20}\text{Ne}/^{22}\text{Ne}$ or Ne/Ar . The present similarity in these two ratios on Mars and Earth therefore indicates a common early history for neon. On Venus we find a larger total abundance of neon as well as a higher value for $^{20}\text{Ne}/^{22}\text{Ne}$, a pattern consistent with mass fractionation by diffusion or atmospheric escape. The normality of the nitrogen isotopes on Venus and Earth argues for atmospheric emplacement of nitrogen after the fractionation of neon on both planets. Cometary impacts could accomplish this. (On Mars the enrichment of ^{15}N can be explained by nonthermal escape^{39,40}.) The near agreement between $^{36}\text{Ar}/^{20}\text{Ne}$ on Venus and that on Earth and Mars must be regarded as a coincidence⁵.

Persuasive as the evidence in Fig. 2b seems, we still need to know whether pristine comets really carry the heavy noble gases in their ices, and if so, in what proportions. In principle, a rocket-borne spectrometer should be able to detect argon in a bright, new comet⁴¹, but a slow flyby or sample return mission is required to measure all of the gases. Are isotopes of Xe and Kr fractionated by low-temperature trapping in ice? It seems unlikely, but this can be tested in the laboratory. Does the unique value of $^{36}\text{Ar}/^{38}\text{Ar} = 4.1 \pm 0.2$ found in EETA79001 glass⁸ really correspond to the martian atmosphere? This value is at the lower limit of the *in situ* measurements of the martian atmosphere by the Viking gas chromatograph-mass spectrometer⁴². If it is the true atmospheric value, we need to ask how it came about. Hydrodynamic escape could fractionate the argon^{3,43} but it would also reduce its total amount, destroying the linearity of the mixing line in Fig. 2b. Finally, we would like to know the relative abundance of Xe on Venus, and the Xe isotope abundances. The answers will determine the location of Venus in Fig. 2b and help to define the source of the planet's volatiles.

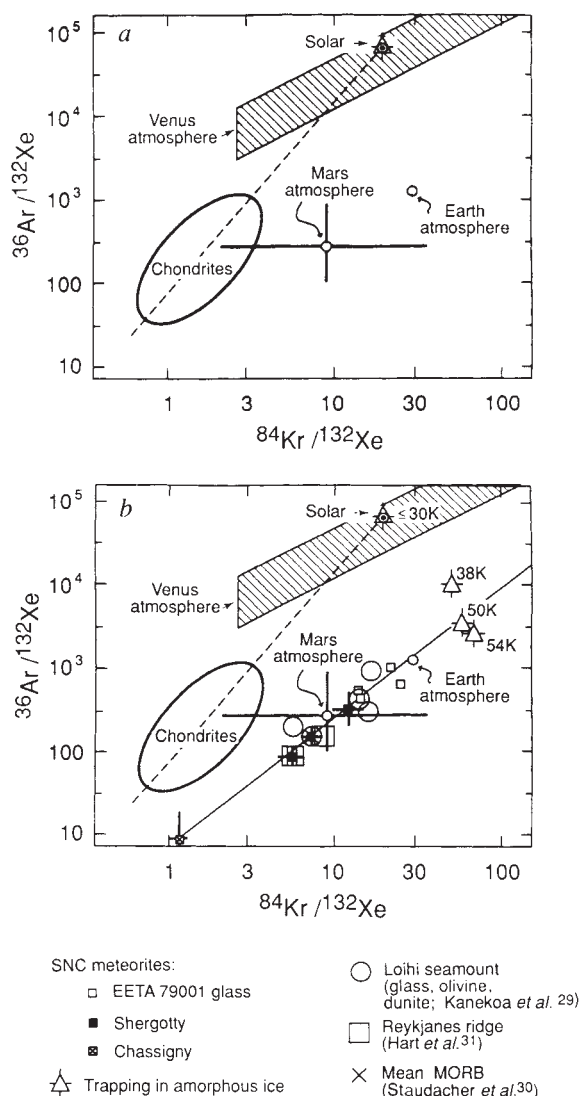


FIG. 2 a, Measurements of argon, krypton and xenon in chondritic meteorites⁹, and the atmospheres of Venus², Mars⁴⁴ and Earth. (The Venus data represent an analysis of *in situ* measurements from instruments carried by Pioneer Venus and Veneras 11 and 12.) Note that two distinct linear families exist in such a log-log plot, one consisting of the chondritic meteorites and the Sun, the other including the atmospheres of Earth and Mars. b, The same as a with the addition of the data obtained from selected SNC meteorites (after Ott and Begemann⁹), noble gas values from mantle-derived terrestrial rocks and the gas mixtures trapped in ice formed at various temperatures in a solar mixture of gases. Note that ice formed at $T \leq 30\text{ K}$ simply reproduces the solar mixture. The MORB point is the mean of several measurements³⁰ and is superimposed on one of the values determined in the Loihi rocks²⁹.

Received 22 January 1991; accepted 12 May 1992.

- Wacker, J. F. & Anders, E. *Geochim. cosmochim. Acta* **48**, 2373-2380 (1984).
- Hoffman, J. H., Oyama, V. I. & von Zahn, U. *J. geophys. Res.* **85**, 7871-7881 (1980).
- Pepin, R. O. *Icarus* **92**, 2-79 (1991).
- Owen, T. in *Workshop on the Evolution of the Martian Atmosphere LPI Tech. Rep. 86-07* (eds Carr, M., James, P., Leovy, C. & Pepin, R.) 31-32 (Lunar and Planetary Science Institute, Houston, 1986).
- Hunten, D. M., Pepin, R. O. & Owen, T. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 565-591 (Univ. of Arizona Press, 1988).
- Owen, T., Bar-Nun, A. & Kleinfeld, I. in *Comets in the Post-Halley Era* (eds Newburn, R. Jr., Neugebauer, M. & Rahe, J.) 429-438 (Kluwer, Dordrecht, 1991).
- Owen, T. in *Mars* (eds Kieffer, H., Jakosky, B., Snyder, C. & Matthews, M.) (Univ. of Arizona Press, in the press).
- Wiens, R. C., Becker, R. H. & Pepin, R. O. *Earth planet. Sci. Lett.* **77**, 149-159 (1986).
- Ott, U. & Begemann, F. *Nature* **317**, 509-512 (1985).
- Ott, U. *Geochim. cosmochim. Acta* **52**, 1937-1948 (1988).
- Laifer, D., Kochavi, E. & Bar-Nun, A. *Phys. Rev. B* **36**, 9219-9227 (1987).
- Bar-Nun, A., Kleinfeld, J. & Kochavi, E. *Phys. Rev. B* **38**, 7749-7754 (1988).
- Wetherill, G. W. in *Mercury* (eds Vilas, F., Chapman, C. R. & Matthews, M. S.) 670-691 (Univ. of Arizona, Tucson, 1988); *A. Rev. Earth planet. Sci.* **18**, 205-256 (1990).

14. Cameron, A. G. W. *Icarus* **56**, 195–201 (1983).
15. Melosh, J. & Vickery, A. *Nature* **338**, 487–489 (1989).
16. Stolper, E. M. & McSweeney, H. Y. *Geochim. cosmochim. Acta* **43**, 1475 (1979).
17. Stolper, E. M., McSweeney, H. Y. & Hays, J. F. *Geochim. cosmochim. Acta* **42**, 589 (1978).
18. Bogard, D. D. & Johnson, P. *Science* **221**, 651–654 (1983).
19. Ott, U., Lohr, H. P. & Begemann, F. *Meteorites* **23**, 295–296 (1988).
20. Watson, L. L., Ihinger, P. D., Epstein, S. & Stolper, E. M. *Proc. lunar planet. Sci. Conf.* (in the press).
21. Musselwhite, D. S., Drake, M. J. & Swindle, T. D. *Nature* **352**, 697–699 (1991).
22. Oro, J. *Nature* **190**, 389–390 (1961).
23. Sill, G. & Wilkening, L. *Icarus* **33**, 13–22 (1978).
24. Chyba, C. *Nature* **330**, 632–635 (1987).
25. Ip, W.-H. & Fernandez, J. A. *Icarus* **74**, 47–61 (1988).
26. Boss, E. G., Morfil, G. E. & Tscharnuter, W. M. in *Origin and Evolution of Planetary Satellite Atmospheres* (eds Atreya, S. K., Pollack, J. B. & Matthews, M. S.) 35 (Univ. of Arizona Press, Tucson, 1989).
27. Lunine, J. I., Engel, S., Rizek, B. & Horanyi, M. *Icarus* **94**, 333 (1991).
28. Bar-Nun, A. & Kleinfeld, I. *Icarus* **80**, 243–253 (1989).
29. Kaneoka, I., Takaoda, N. & Clague, D. A. *Earth planet. Sci. Lett.* **66**, 427–437 (1983).
30. Studacher, T., Kunz, M. D. & Allegre, C. J. *Chem. Geol.* **56**, 193–205 (1986).
31. Hart, R., Dymond, J., Hogan, L. & Schilling, J. G. *Nature* **305**, 403 (1983).
32. Wetherill, G. W. *Icarus* **46**, 70–80 (1981).
33. Clayton, R. N. & Mayeda, T. K. *Earth planet. Sci. Lett.* **62**, 1–6 (1983).
34. Meech, K. J. & Belton, M. J. S. *IAU Circ. No.* 4770, 11 April 1989.
35. Luu, J. X. & Jewitt, D. C. *Astr. J.* **100**, 913–932 (1990).
36. Bus, S. J., A'Hearn, M. F., Schleicher, D. G. & Bowell, F. *Science* **251**, 774–777 (1991).
37. Craig, H. & Lupton, J. E. *Earth planet. Sci. Lett.* **31**, 369–385 (1987).
38. Honda, M., McDougall, I., Patterson, D. B., Doulgeris, A. & Clague, D. A. *Nature* **349**, 149–151 (1991).
39. McElroy, M. B. *Science* **175**, 443–445 (1972).
40. Fox, J. in *Dissociative Recombination: Theory, Experiment and Applications* (eds Mitchell, J. B. A. & Guberman, F. L.) 264–281 (World Scientific, Singapore, 1989).
41. Stern, S. A., Green, J. C., Cash, W. & Cook, T. A. *Icarus* **95**, 157–161 (1992).
42. Biemann, K., Owen, T., Rushneck, D. R., LaFleur, A. L. & Howarth, D. W. *Science* **194**, 76–78 (1976).
43. Zahnle, K., Kasting, J. F. & Pollack, J. B. *Icarus* **84**, 502–527 (1990).
44. Owen, T. *et al.* *J. geophys. Res.* **82**, 4635–4639 (1977).

ACKNOWLEDGEMENTS. This research was supported in part by NASA.

A high-pressure van der Waals compound in solid nitrogen–helium mixtures

W. L. Vos, L. W. Finger, R. J. Hemley, J. Z. Hu, H. K. Mao & J. A. Schouten*

Geophysical Laboratory and Center for High Pressure Research, Carnegie Institution of Washington, 5251 Broad Branch Road NW, Washington DC 20015, USA

* van der Waals-Zeeman Laboratorium, Universiteit van Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

WEAKLY bound van der Waals molecules, consisting of atoms or molecules (such as rare gases, H_2 and N_2) interacting through van der Waals forces, can be formed in the gas phase¹. The existence of similar weakly bound stoichiometric compounds in condensed phases has remained an open question as the components usually form separate pure phases or solid solutions². Here we report on a detailed study of the helium–nitrogen system in a diamond-anvil cell using synchrotron X-ray diffraction, Raman scattering and optical microscopy. We find that high pressure stabilizes the formation of a stoichiometric, solid van der Waals compound of composition $He(N_2)_{11}$. Because of the relatively simple interactions involved, this solid may exemplify a novel class of van der Waals compounds that are formed only at high pressures. Van der Waals chemistry in simple molecular systems may be important for understanding the structure and properties of the interiors of the outer planets and their satellites, where pressures are high and such components may be abundant^{3–5}.

Simple binary mixtures such as He– H_2 , He–Ne, He– H_2 and N_2 – O_2 , compressed at pressures up to ~20 GPa in diamond cells⁶, typically show limited mutual solubility in the solid phase⁷. Helium has an extraordinary effect on the phase behaviour of solid N_2 . At high pressure and room temperature pure nitrogen has two distinct polymorphs: the δ -phase, a cubic phase with orientational disorder⁸, and the ϵ -phase, a rhombohedral orientationally ordered phase⁹ which occurs above 16.3 GPa (ref. 10). Pure helium freezes at 12.0 GPa into a hexagonally close-packed structure^{11,12}. For He– N_2 mixtures,

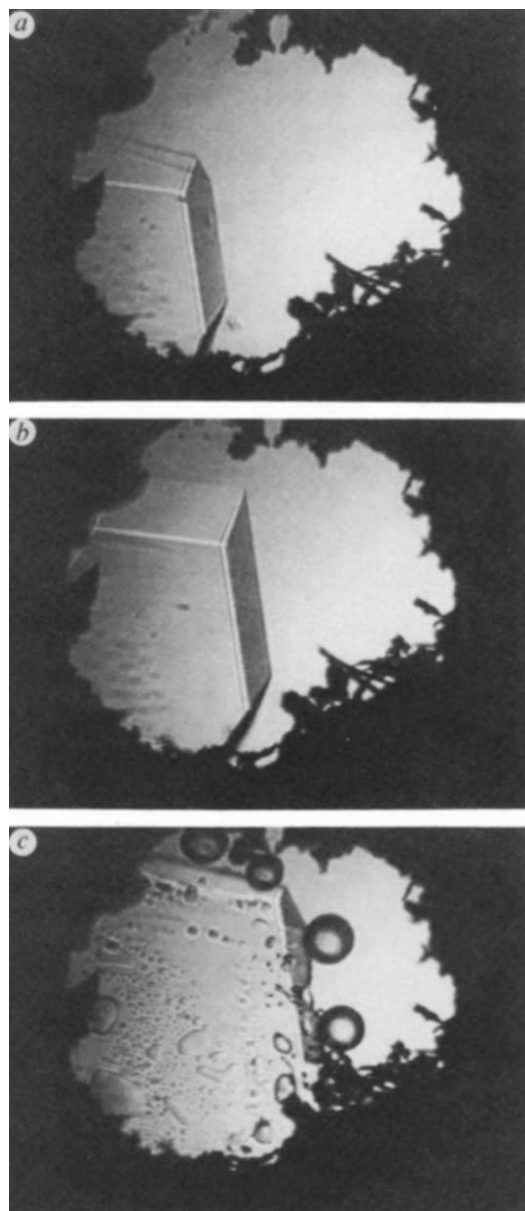


FIG. 1 Photomicrographs of the growth of a single-crystal of the high-pressure He– N_2 phase. The bulk composition is 41.6 mol% He and the pressure ~8.9 GPa. *a*, Temperature of 315 K. Unlike other high-pressure van der Waals crystals, facets are clearly seen during the growth. *b*, 308 K. One of the growth directions seen in *a* has disappeared against the face of one of the diamond anvils. *c*, Three-phase equilibrium solid–fluid–fluid¹⁵ at 305 K, where the N_2 -rich fluid decomposes into the solid phase and a He-rich fluid (visible as bubbles). The latter phase blurs the facets and introduces rounded gas inclusions in the crystal. The experiments were done on gases that were mixed from high-purity components (99.999%). High pressure was generated with single-crystal Mao–Bell diamond-anvil cells²⁴. The samples were loaded at pressures of ~0.2 GPa (refs 12, 17). The pressures were measured with the ruby fluorescence scale and are expected to be accurate to ~0.2 GPa.

about 2 mol% He can be dissolved in orientationally disordered phases at high pressures and low temperatures^{13–15}, and it seems that about 9 mol% can be dissolved in the ϵ -phase at higher pressure^{13–15}. This remarkably high solubility results in a marked shift of the orientational order–disorder transition by a factor of two upwards in temperature or downwards in pressure¹³. In previous work on binary mixtures, however, no direct structural

information was available on the high-pressure phases, including the '9% phase' of the He-N₂ system. This information is essential for proper phase identification and is obtained here by the use of new high-pressure single-crystal techniques.

A sequence of photomicrographs of the '9% phase' in a He-N₂ mixture *in situ* in the diamond-anvil cell at a pressure near 9 GPa is shown in Fig. 1. A single crystal of the new phase is seen to grow within the surrounding He-rich fluid as the temperature decreases from 315 to 305 K. The crystallinity of this phase is clearly demonstrated by its morphology. Well developed facets are seen during the growth, whereas van der Waals crystals of pure components (such as He and H₂) growing at high pressure develop rounded surfaces^{12,15-17}. The crystal is birefringent under crossed polarizers and can thus be distinguished from the isotropic δ -phase of solid N₂. The lower image (305 K) shows the decomposition of the N₂-rich fluid into the solid phase and a He-rich fluid, which is visible as bubbles.

In Fig. 2, we show X-ray diffraction patterns measured at 9.2 GPa and 295 K on a single crystal formed from a bulk composition of 10.0 mol% He. The reflections had sharp rocking curves of about 0.1° in ω , indicative of high-quality crystals. From the 21 observed d -spacings, a hexagonal unit cell was obtained. The reflections of this phase do not fulfil the condition $-h+k+l=3n$ (n is an integer) of rhombohedral structures. These results indicate that the '9% phase' is a new phase with a structure distinct from any of the known high-pressure phases of pure nitrogen and helium.

We looked for additional phases in mixtures of different compositions (5.0, 19.6 and 41.6 mol% He). The d -spacings obtained for measurements on single crystals of the phase grown from a 5.0 mol% He composition are listed in Table 1. All new peaks could be fitted to the hexagonal unit cell of the new phase, hereafter designated the van der Waals (vdW) compound. Two additional lines fitted the strong 210 and 211 reflections of the δ -phase^{8,10}. The 103 reflection overlaps with the 200 reflection of the δ -phase.

We used *in situ* Raman spectroscopy to obtain additional information on the new phase. Measurements on the vdW com-

TABLE 1 Calculated and observed lattice spacings at 12.6 GPa and bulk composition 5.0 mol% He

d_{obs}	vdW compound		δ -phase	
	$h\ k\ l$	d_{calc}	$h\ k\ l$	d_{calc}
2.868	1 0 3	2.871	2 0 0	2.868
2.801	2 0 2	2.799		
2.631	2 1 0	2.625		
2.565			2 1 0	2.565
2.533	2 1 1	2.529		
2.479	1 1 3	2.479		
2.341			2 1 1	2.341
2.299	2 1 2	2.295		
2.241	3 0 1	2.249		

The unit cell of the hexagonal vdW compound has dimensions $a=8.019(5)$ Å and $c=9.461(12)$ Å. The lattice parameter of the cubic δ -phase is $a=5.735(1)$ Å.

pounds with compositions of 10.0, 19.6 and 41.6 mol% He revealed a main vibron shifted with respect to that of pure N₂ (for example $\Delta\nu = 1.5\text{ cm}^{-1}$ at $P = 14$ GPa). With 5.0 mol% He, the main vibrons of both the vdW compound and the N₂ phases were observed. We found that the measured relative intensities of the two peaks varied if we pointed the laser beam at different spots in the sample. Because the intensities from the δ - and ϵ -phases are large in the X-ray diffraction and the Raman scattering experiments, these phases must be a sizeable portion of the volume. Therefore, from mass-balance considerations, the composition of the new vdW compound phase is greater than 5.0 mol% He. On the other hand, it must be smaller than 10.0 mol%, because experiments at 10.0, 19.6 and 41.6 mol% He showed identical spectra. At 7.7 GPa for the 10 mol% composition, the δ -phase and the vdW compound coexisted, so we take this as the formation pressure of the compound at room temperature.

To investigate whether the vdW compound is stoichiometric, we compared the unit-cell volumes of this phase, measured at different compositions but at the same pressure. If the phase is stoichiometric, the unit-cell volume does not vary with the initial bulk composition, whereas if it is non-stoichiometric, the volume will depend on composition because the constituent species differ in size. The variation should increase with pressure starting from the formation pressure. The differences can be best observed by taking compositions below the lowest composition (we use the results at 5.0 mol%) and above the highest composition of the vdW compound (results at 10.0 and 41.6 mol%). The volumes at 41.6 mol% were obtained from the 103, 104 and 004 reflections (see Fig. 3). The results for 10.0 and 41.6 mol% are very consistent, in agreement with the upper composition bound of 10.0 mol% He. Of the two points obtained at 5.0 mol%, the one at lower pressure is slightly higher, but the second is indistinguishable from the other results. We conclude that there is no composition effect on the volume of the vdW compound and that it is therefore stoichiometric.

The contents of the unit cell were constrained on the basis of the thermodynamic information. The Gibbs free enthalpy difference between the compound phase and the coexisting phases was calculated from $\Delta G = \int \Delta V(p) dp$ (ΔV is the volume difference between the vdW compound and the phases with which it is in equilibrium) starting from the formation pressure at 7.7 GPa where the enthalpy difference is zero. We used known $V(p)$ isotherms of N₂ and He (refs 8, 10, 12, 18) and known compositions of the fluid phases¹⁵. The positive volume change on mixing for the fluid phases was assumed to increase parabolically to 2% at the critical composition as indicated by theoretical results for He-H₂ (ref. 19). This analysis indicates that the vdW compound is stable if it contains at least 22 N₂ molecules and 2 He atoms per unit cell, giving a hexagonal unit cell with 24

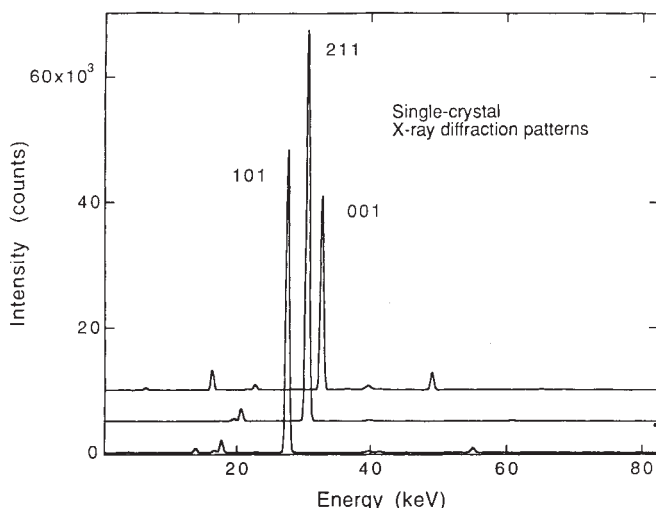


FIG. 2 Diffraction patterns of the vdW compound with 10 mol% helium taken at $2\theta = 9^\circ$. Reflections of the 101, 211 and 001 classes are shown. The 211 and 001 spectra are offset for clarity. The large peaks of the 101 and 001 classes are the second and fourth orders. Careful inspection revealed up to four orders in these reflections. Small peaks 10 keV below the main reflections are escape peaks from the germanium detector. As the low-Z elements and the very small samples involved require a high-brightness source, X-ray diffraction studies were done at the superconducting wiggler beamline X-17C of the National Synchrotron Light Source, at Brookhaven National Laboratory²⁵. The accuracy of the unit-cell volumes is estimated as 0.2–0.3%.

particles. This stoichiometry is in excellent agreement with the composition bounds (5.0–10.0%) deduced above. The volume difference from pure solid nitrogen and helium (dashed line in Fig. 3) is reasonable in view of the pressure jumps measured previously at transitions involving this phase¹³. This is also consistent with the fact that the Raman shifts of this phase are similar to those of the δ - and ϵ -phases at the same pressure.

The stability of the new compound can be understood in part on the basis of hard-sphere models, as used to explain mixing behaviour in simple systems²⁰. Mixtures of spherical particles with a diameter ratio between 1.0 and ~ 0.85 form random substitutional solid solutions. In very dissimilar systems with size ratios up to about 0.57, small particles can occupy the interstices between the larger ones. In both cases, the composition is continuously variable. Systems with intermediate size ratios separate into pure phases. Alternatively, Murray and Sanders have argued that for size ratios up to 0.62, ordered structures AB_2 and AB_{13} may be stable²¹. The size ratio for N_2 and He is roughly 0.6–0.7. Our experiments demonstrate that stoichiometric compounds form in molecular-atomic systems at high pressure. The driving force for the formation of the phase observed here is its higher density relative to that of the endmember assemblage (Fig. 3). Previous high-pressure studies have, in general, been done on binary systems with size ratios larger than about 0.8, and this may be why such compounds have not been observed. Important differences between the present observations and previous hard-sphere results are the unusual stoichiometry²² (AB_{11}) and crystal structure of the compound, and the fact that the large species is the major component. The more-complex behaviour of the molecular system is probably associated with pressure dependence of the effective radii, or with anisotropy and non-additivity of the intermolecular potentials. On the basis of size ratios, compounds can be expected in the binary atomic systems He–Ar and He–Xe. The mixture H_2 –Xe may show a change in bonding from the new high-pressure van der Waals chemistry identified here to metallic bonding at very high pressure²³.

Finally, we suggest that the formation of a novel high-pressure N_2 –He phase is probably not unique, but could represent a new class of high-pressure compounds in formally closed-shell van der Waals systems. The unusually large stoichiometric ratio for $He(N_2)_{11}$ suggests that a minor component can strongly affect phase stability in van der Waals systems by forming structures

that are distinct from those of either endmember. In planetary interiors, especially those of the outer planets, molecular species characterized by van der Waals interactions predominate at very high pressures^{3–5}. Knowledge of the compound phases could alter current planetary models based on phase diagrams that assume mixtures of endmembers and solid-solution phases typical of low-pressure behaviour^{3–5}. Further high-pressure studies of multicomponent van der Waals systems are needed to examine these possibilities. □

Received 6 March; accepted 8 May 1992.

- Novick, S. E., Leopold, K. R. & Klempner, W. in *Atomic and Molecular Clusters* (ed. Bernstein, E. R.) 359–391 (Elsevier, New York, 1990).
- Meyer, L. *Adv. chem. Phys.* **16**, 343–387 (1969).
- Hubbard, W. B. *Planetary Interiors* (van Nostrand, New York, 1984).
- Nellis, W. J. *et al. Science* **240**, 779–781 (1988).
- Hubbard, W. B. *et al. Science* **253**, 648–651 (1991).
- Schouten, J. A. *Phys. Rep.* **172**, 33–92 (1989).
- Schouten, J. A. *Fluid Phase Equil.* (in the press).
- Cromer, D. T., Mills, R. L., Schiferl, D. & Schwalbe, L. A. *Acta crystallogr.* **B37**, 8–11 (1981).
- Mills, R. L., Cromer, D. T. & Olinger, B. J. *chem. Phys.* **84**, 2837–2845 (1986).
- Olijnyk, H. J. *chem. Phys.* **93**, 8968–8972 (1990).
- Vos, W. L., van Hinsberg, M. G. E. & Schouten, J. A. *Phys. Rev.* **B42**, 6106–6109 (1990).
- Mao, H. K. *et al. Phys. Rev. Lett.* **60**, 2649–2652 (1988).
- Vos, W. L. & Schouten, J. A. *Phys. Rev. Lett.* **64**, 898–901 (1990).
- Scheerboom, M. I. M. & Schouten, J. A. *J. Phys.: Condensed Matter* **3**, 8305–8312 (1991).
- Vos, W. L. & Schouten, J. A. *Physica A* **182**, 365–387 (1992).
- van den Bergh, L. C., Schouten, J. A. & Trappeniers, N. J. *Physica A* **141**, 524–538 (1987).
- Mao, H. K. *et al. Science* **239**, 1131–1134 (1988).
- Le Toullec, R., Loubeyre, P. & Pinceaux, J.-P. *Phys. Rev.* **B40**, 2368–2378 (1990).
- van den Bergh, L. C. & Schouten, J. A. *J. chem. Phys.* **89**, 2336–2343 (1988).
- Bartlett, P. J. *J. Phys.: Condensed Matter* **2**, 4979–4989 (1990).
- Murray, M. J. & Sanders, J. V. *Phil. Mag.* **A42**, 721–740 (1980).
- Kitaigorodsky, A. I. *Mixed Crystals* (Springer, Berlin, 1984).
- Hemley, R. J. & Mao, H. K. in *Proc. APS 1991 Topical Conf. Shock Compression of Condensed Matter* (eds Schmidt, S. C., Dick, R. D., Forbes, J. W. & Tasker, D. G.) (in the press).
- Mao, H. K. in *Simple Molecular Systems at Very High Densities* (eds Polian, A., Loubeyre, P. & Boccara, N.) 221–236 (Plenum, New York, 1989).
- Hu, J. Z., Mao, H. K. & Shu, J. F. *EOS* **72**, 554 (1991).

ACKNOWLEDGEMENTS. We thank M. Scheerboom, J.-L. Barrat, D. Frenkel, A. de Kijper, R. Le Sar, and C. Meade for discussions, and M. Hanfland and J. F. Shu for experimental help. W.L.V. is also grateful to R. Sprk and R. Kroon for help with Raman experiments. This work was supported by NASA, NATO, NSF and the Carnegie Institution of Washington.

Densification of nanostructured titania assisted by a phase transformation

Krishnankutty-Nair P. Kumar*, K. Keizer*,
A. J. Burggraaf*, Tatsuya Okubo†,
Hidetoshi Nagamoto† & Shigeharu Morooka‡

* Laboratory for Inorganic Chemistry, Materials Science and Catalysis,
Faculty of Chemical Technology, University of Twente,
7500 AE Enschede, The Netherlands

† Engineering Research Institute, Faculty of Engineering,
The University of Tokyo, Tokyo 113, Japan

‡ Department of Chemical Science and Technology, Faculty of Engineering,
Kyushu University, Fukuoka 812, Japan

NANOPHASE materials, characterized by an ultrafine grain size, have stimulated much interest in recent years^{1–11} by virtue of their unusual mechanical, electrical, optical and magnetic properties. Nanophase ceramics are of particular interest because they are more ductile at elevated temperatures than are coarse-grained ceramics¹¹—an important property for the fabrication of ceramic components. Preparing materials that are both dense and fine-grained, however, has proved difficult: the high sintering temperatures generally required to obtain high densities can also lead to exaggerated grain growth, resulting in coarse-grained, nonuniform materials. Sintering at lower temperatures gives a much finer grain size, but does not in general result in high-density materials. We show here that dense nanostructured titania, with density >99% of the theoretical maximum and an average grain size of less than 60 nm, can be prepared by sintering a titanium oxide

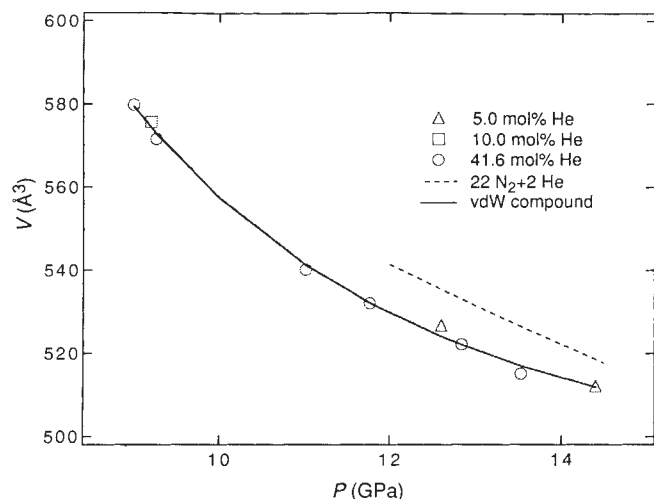


FIG. 3 Unit-cell volume, V , of the vdW compound as a function of pressure P obtained at compositions of 5.0, 10.0 and 41.6 mol% He. The dotted line shows the pressure–volume relations determined for pure nitrogen (volume per 22 molecules)¹⁰. The dashed curve shows the pressure–volume relation for a hypothetical solid mixture of 22 molecules of nitrogen¹⁰ and two of helium^{12,18} assuming no high-pressure van der Waals reaction.

sol-gel near the anatase-rutile phase transformation temperature (about 600 °C). The increased mobility of the atoms during the phase transformation enhances the sintering rate at lower temperatures, suggesting that this method could be used more generally to produce nanophase materials with near theoretical densities.

Nanophase metals and ceramic powders are normally synthesized by evaporating the metal in an inert atmosphere and allowing it to condense on the surface of a 'cold finger' kept at liquid nitrogen temperature¹. To produce fine ceramic powders, the fine metal clusters are oxidized and polycrystalline ceramic samples produced by *in situ* compaction inside the reaction chamber^{1,2}. Titanium dioxide produced in this way can be sintered to ~96–99% of its theoretical density at 1,000 °C with a grain size of 1,000 nm (ref. 7). Sintering at a lower temperature (900 °C) yielded a much finer grain size (~51 nm) but only 91% of the theoretical density⁷. Hahn *et al.*⁷ obtained dense titania with a grain size of 40 nm by pressure-assisted sintering and sinter-forging with near theoretical densities.

The sol-gel technique is an alternative route to the production of nanocrystalline materials, allowing fine powders (3 to 10 nm primary crystallite size⁴) with a very high level of chemical purity to be obtained. But there are problems such as aggregation, contamination during processing and residual small pores remain. Gel-derived TiO₂ powder produced by the hydrolysis of titanium tetra isopropoxide can be sintered to almost 100% of the theoretical density at 800 °C with a fine grain size of ~150 nm (ref. 8). This was achieved by carefully controlling the

processing variables to avoid aggregation and to obtain good uniform packing in the green compacted stage. The particle size distribution was also reported to be narrow. The sintered microstructure is largely determined by the powder characteristics and the initial microstructure^{8,12}.

We prepared a titanium oxide sol by hydrolysing (at 25 °C) alcoholic solutions of titanium isopropoxide (0.5 molar solution in isopropyl alcohol) with an alcoholic solution (9% water in isopropyl alcohol) of water. The molar ratio of water to alkoxide was 20. Reaction conditions were carefully controlled to give a titanium oxyhydroxide precipitate, which was washed with water and then redispersed in water. A portion of this dispersion was kept for further analysis and comparison. The other portion was peptized by adding HNO₃, followed by refluxing at 80 °C to give a sol of pH ~2. The titania hydrosol thus prepared, which had a concentration of 0.3 mol, was gelled by drying at 40 °C and 60% relative humidity, then calcined to give dense *n*-TiO₂. A typical sample size was 20 × 10 × 0.1 mm. Figure 1a and b shows the X-ray diffraction patterns of the dried gel, both peptized and unpeptized, and the samples calcined at different temperatures. All the peaks for both peptized and unpeptized samples (Fig. 1a) correspond to the pure anatase phase. The very broad peaks indicate the ultra-fine primary particle size. From analysis of the X-ray line broadening, using the Scherrer formula, we calculated a primary particle size of ~6 nm for the peptized sample. This value is in good agreement with our results from small-angle X-ray scattering studies. The only difference we can observe between the peptized and the unpeptized samples is that the peptized one is more crystalline, as indicated by the relatively narrower peaks. Other reported gel-derived titania powders are amorphous in the as-prepared state^{8,10}.

In this processing route, the peptization and drying steps are very important. Peptization will give rise to a sol with a low degree of aggregation. The electrostatically stabilized primary particles pack well during gelling and produce a uniform initial (green) microstructure. This aspect is explained later. During drying, because of the capillary pull of the pore fluid, any aggregates will break down, and a uniform close-packed structure results. This way of processing has many advantages, the most important of which are: (1) the particles will not be exposed to the atmosphere, thus avoiding contamination during the processing; (2) a stabilized sol will give fewer aggregated particles than any other type of starting slurry; (3) the particle packing and rearrangement during drying of the peptized gel is excellent, especially when the pore fluid is water which has a high surface tension. Barringer *et al.*⁹ noted that liquid-phase sintering and hot pressing, which can be used to achieve near theoretical density, essentially modify the initial microstructure in the same way. In the first stage of liquid-phase sintering, the sharp edges of the particles are dissolved and the particles are rearranged by the surface tension of the liquid⁹. In hot pressing it is the applied pressure that causes the rearrangement to give a uniform close-packed structure⁹. We measured isotherms for adsorption of N₂ on the peptized and the unpeptized gels. That for the peptized gel had no hysteresis loop and was a good example of type I, typical of microporous materials (pore diameter less than 2 nm; IUPAC definition). The absence of hysteresis (desorption branch follows the same path as the adsorption branch) clearly indicated the absence of mesopores (no capillary condensation). The isotherm corresponding to the unpeptized gel was a typical type IV isotherm with two hysteresis loops, indicating a bimodal pore-size distribution. These results show that peptization followed by drying is important in obtaining a uniform, well-compacted initial structure. From the amount adsorbed we can see that the unpeptized samples have a higher initial porosity, which is unfavourable for sintering to near theoretical densities.

Figure 2a shows the uniform distribution of pores. The total porosity estimated from the photomicrograph is ~40%; the value obtained from N₂ adsorption is ~30%. During drying, the primary sol particles will rearrange and will be compacted to

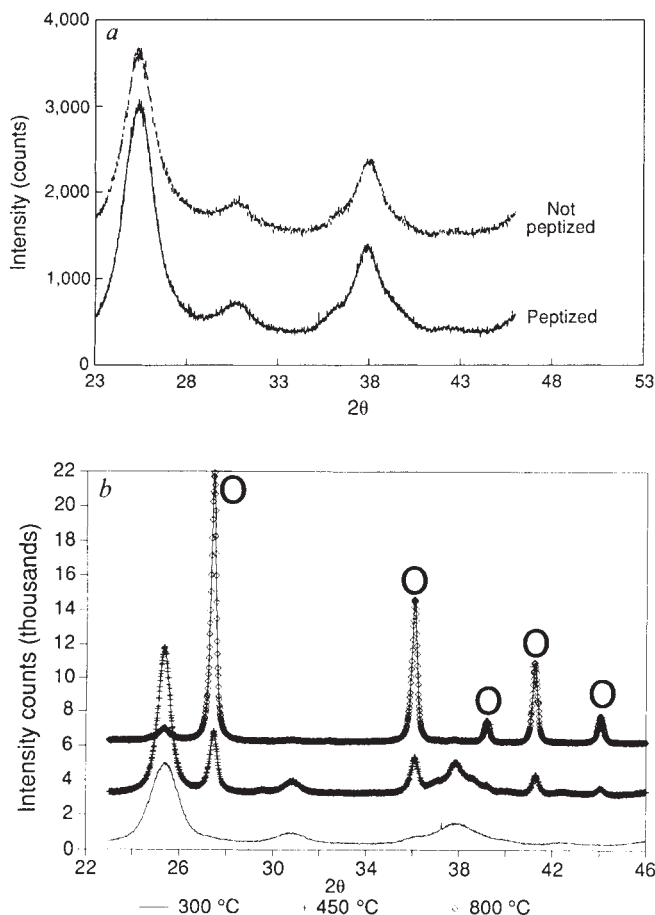
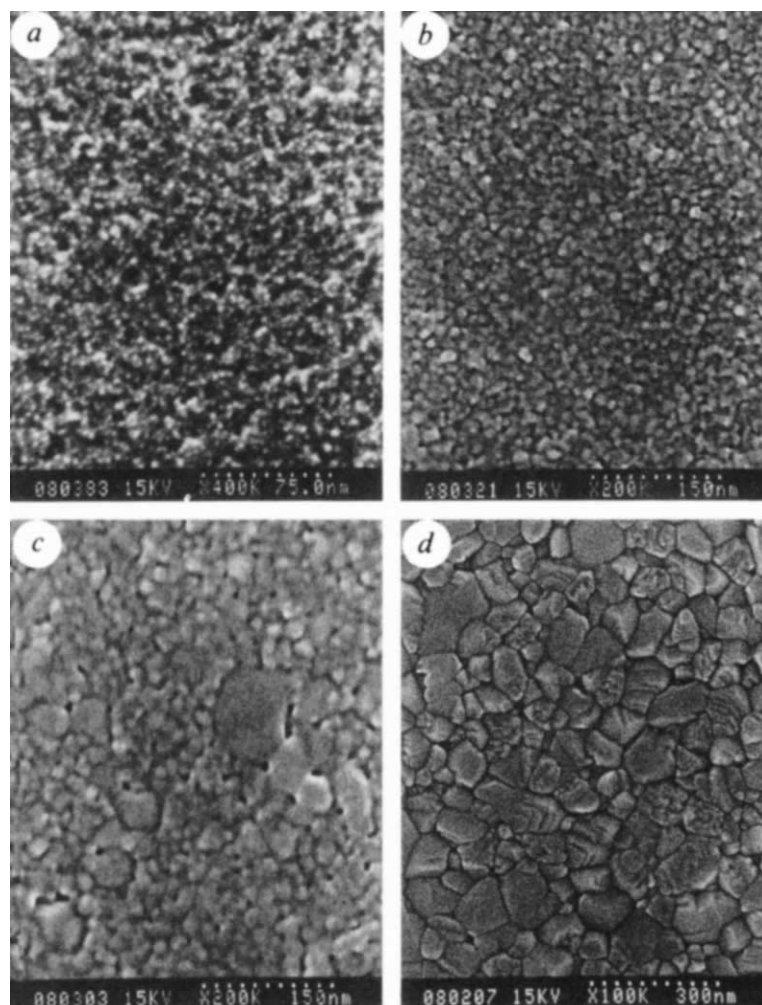


FIG. 1 X-ray diffraction patterns of the titania sample. a, Precursor gel, peptized and unpeptized, heated at 40 °C and 60% relative humidity; b, samples heated at 300, 450 and 600 °C for eight hours. The peaks marked 'O' correspond to the rutile reflections. At 300 °C the phase is essentially anatase.

FIG. 2 High-resolution field-emission scanning electron micrograph of nanostructured titania samples. *a*, Heated at 40 °C and 60% relative humidity (dried gel). The small bright dots are from the platinum coating and the dark parts represent the pores. *b*, Heated at 300 °C. The topology remains similar to the dried gel. *c*, Heated at 450 °C. The porosity is ~20%, anatase grains are roughly 15–20 nm across, and the large grains represent the rutile phase (Fig. 1*b*). *d*, Heated at 600 °C. The microscope magnifications are 400,000, 200,000, 200,000 and 100,000 respectively. The actual magnification can be seen from the bar size.



give a favourable initial microstructure. The topology of the sample heated to 300 °C remains similar to that of the dried gel. At 450 °C (Fig. 1*b*), new peaks corresponding to rutile start to appear. At 600 °C, the structure is essentially rutile with minor amounts of anatase.

Figure 3 shows the results of differential scanning calorimetry of the titania gel, with a heating rate of 12 °C min⁻¹, which is essentially anatase in the as-prepared state. The exothermic peak at ~560 °C represents the discontinuous grain growth of the transformed rutile particles. This has been confirmed by measuring the size of rutile grains in samples containing different amounts of rutile¹³. The exotherm is slightly asymmetric, with the larger area to the left, indicating that heat release started before actual grain growth. The initial heat release is attributed to the enthalpy of the anatase–rutile transformation. This type of asymmetric exotherm is typical of grain growth initiated by phase transformation¹⁴. The total heat change associated with the exotherm was calculated to be 11.4 ± 2 kJ mol⁻¹. The enthalpy of anatase–rutile transformation is only 6.51 kJ mol⁻¹ (ref. 15), so in this case the excess 4.9 kJ mol⁻¹ is the heat released because of grain growth.

Figure 4 shows the results of thermogravimetric analysis of the titania sample at a heating rate of 10 °C min⁻¹. There are two regions of weight loss: the one below 150 °C is due to the loss of physically adsorbed water from the pores, and that in the region 150–400 °C is due to dehydroxylation and the removal of residual organics. There is no weight loss at temperatures above 400 °C. The surface energy associated with grain growth can be calculated from the change in surface area. We calculated

crystallite size from X-ray line-broadening studies, and found a change from 78 m² g⁻¹ before transformation and crystal growth to 28 m² g⁻¹ after the process, or ~50 m² g⁻¹. From this and the enthalpy for grain growth (4.8 kJ mol⁻¹, 60.8 J g⁻¹) we calculated that the surface energy released was 1.1 ± 0.1 J m⁻². This relatively high energy release¹⁶ may contribute to the increased rate of sintering at lower temperatures by causing a temperature rise. At 600 °C, (Fig. 2*d*) the sample has converted completely to rutile at 99% of the theoretical density. The average grain size was calculated to be ~60 nm. During transformation, atoms are very mobile because of the bond breakage, and an enhanced sintering is expected near this temperature. In practice this is not always true, sintering of alumina gel being a typical

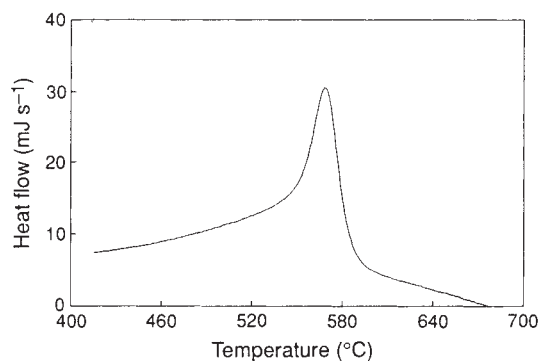


FIG. 3 Differential scanning calorimeter trace of titania gel.

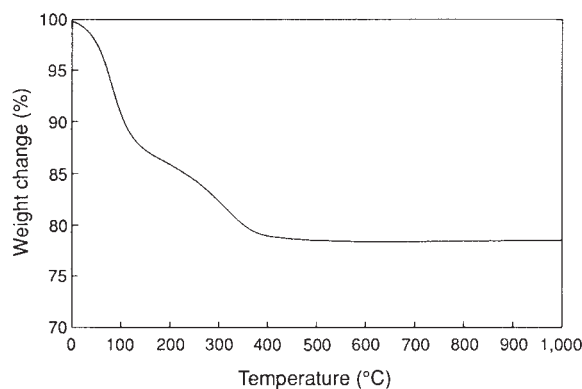


FIG. 4 Thermogravimetry trace of titania gel.

example. The most important difference between the two transformations (*trans*-alumina to α -alumina and anatase to rutile) is that the nucleation rate of alumina is very low compared with the growth rate. For titania (especially that prepared by the technique described here) the rate of nucleation is comparable to that of growth. In fact if we manipulate the precursor chemistry by adjusting the hydrolytic polycondensation parameters or by other chemical modifications, and so increase the

rate of nucleation during transformation, we may obtain dense materials with even smaller grain sizes. One nucleus per primary particle of anatase is the ideal situation, and we are investigating this possibility. □

Received 19 November 1991; accepted 14 May 1992.

1. Birringer, R., Herr, U. & Gleiter, H. *Trans. Jap. Inst. Metall. Suppl.* **27**, 43–45 (1986).
2. Karch, J., Birringer, R. & Gleiter, H. *Nature* **330**, 556–558 (1987).
3. Siegel, R. W. & Eastman, J. A. in *Multicomponent Ultrafine Microstructures*, MRS Symp. Proc. Vol. 132, 3–14 (1988).
4. Xu, Q. & Anderson, M. A. in *Multicomponent Ultrafine Microstructures*, MRS Symp. Proc. Vol. 132, 41–46 (1988).
5. Siegel, R. W. in *Superplasticity in Metals, Ceramics, and Intermetallics*, MRS Symp. Proc. Vol. 196, 59–70 (1990).
6. Boutz, M. M. R., Theunissen, G. S. A. M., Winnubst, A. J. A. & Burggraaf, A. J. in *Superplasticity in Metals, Ceramics, and Intermetallics*, MRS Symp. Proc. Vol. 196, 87–92 (1990).
7. Hahn, H., Logas, J., Hofler, H. J., Kurath, P. & Averbach, R. S. in *Superplasticity in Metals, Ceramics, and Intermetallics*, MRS Symp. Proc. Vol. 196, 71–76 (1990).
8. Barringer, E. A. & Bowen, H. K. *J. Am. Ceram. Soc.* **65**, C199–C201 (1982).
9. Barringer, E. A., Jubb, N., Fegley, B., Pober, R. L. & Bowen, H. K. in *Ultrastructure Processing of Ceramics, Glasses, and Composites* (eds Hench, L. L. & Ulrich, D. R.) 315–333 (1984).
10. Springer, L. & Yan, M. F. in *Ultrastructure Processing of Ceramics, Glasses, and Composites* (eds Hench, L. L. & Ulrich, D. R.) 464–475 (1984).
11. Nieh, T. G., Wadsworth, J. & Wakai, F. *Int. Mat. Rev.* **36**, 146–161 (1991).
12. Onoda, Jr G. Y. & Hench, L. L. *Ceramic Processing Before Firing* (Wiley, New York, 1978).
13. Kumar, K. P. thesis, Univ. of Twente (1993).
14. Sheinkman, A. I., Tyumentsev, V. A. & Fotiev, A. A. *Izv. Akad. Nauk SSSR, Neorg. Mater.* **20**, 1692–1696 (1984).
15. JANAF Thermochemical Tables, ACS, AIP, NBS 3rd Edn. (eds Chase et al.) (1985).
16. Lu, K., Wang, J. T. & Wei, W. D. *J. appl. Phys.* **69**, 522 (1991).

ACKNOWLEDGEMENTS. K.-N.P.K. would like to dedicate this work to the memory of Sri. N. C. Pillai. We thank M. M. R. Boutz and Peter den Exter for discussion.

Possible connection between surface winds, solar activity and the Earth's magnetic field

Roger Y. Anderson

Department of Geology, University of New Mexico, Albuquerque, New Mexico 87131, USA

RECORDS of ^{14}C in tree rings provide a proxy for changes in solar activity on timescales of decades to centuries^{1,2}. Here I report an association between Maunder-type solar cycles recorded in tree rings and fluctuations in surface wind intensity on a roughly 200-year timescale over a period of about 2,000 years in the mid-Holocene. The wind record is preserved as changes in the thickness of varved sediments in Elk Lake, Minnesota, which contain a large aeolian component. Changes in cyclonic activity and tropospheric winds have been reported^{3–5} to occur in this region a few days after strong coronal mass ejections (CMEs); increases of up to 7% in zonal flow have been reported⁵ at an altitude of ~300 m after strong CMEs in winter. The implication of this association is that century-scale changes in surface winds over Minnesota were the long-term (Maunder-scale) counterpart of short-term (daily) changes in winds after CMEs. The weak magnetic field strength of the Earth at this time^{6,7} might be relevant to a possible mechanism for the association.

Varved sediments began accumulating in Elk Lake (47° 13' N, 95° W) ~10,400 years ago⁸. Before 8,400 years, the varves contain little or no aeolian material and are less than 1 mm thick (Fig. 1a), reflecting a stable forest cover in the area around the lake. Prairie conditions moved eastward over northern Minnesota about 8,400 years ago, replacing forest with areas of bare ground and scattered dune fields⁹. Elk Lake lies ~70 km southeast and downwind from the margin of glacial Lake Agassiz, and some of the fine-grained sediment (loess) that accumulated in Elk Lake was probably derived from exposed Late Pleistocene lake beds. Between ~8,400 and ~3,800 years ago, aeolian clay and silt was carried into Elk Lake (Fig. 1b) and resulted in a threefold increase in varve thickness, as well as a marked increase in variance. Changes in varve thickness

therefore define changes in palaeowinds. The varve time series from Elk Lake contains a ~600-year interval, centred at ~5,000 years, during which the accumulation of clay and silt was abruptly terminated and then restarted (Fig. 1a). This mid-Holocene event has been identified in aeolian sediments in Lake Ann¹⁰, 200 km southeast of Elk Lake, and suggests that palaeowind activity over Minnesota, as recorded in Elk Lake, was regional in extent.

The 1-km² basin that holds Elk Lake contains 30 m of water, underlain by 22 m of varved sediments which are varved to the

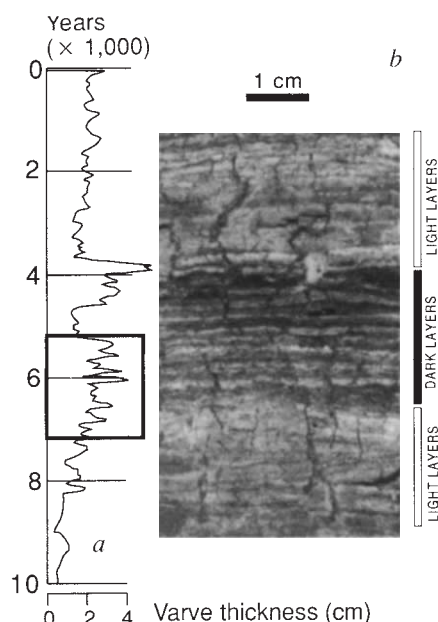


FIG. 1 Filtered time series for varve thickness at Elk Lake, and photograph of thick varves formed during the mid-Holocene. a, Oscillations in varve thickness of ~200 years, 40–50 years and ~22 years occur in the indicated time-window; b, Thicker, lighter-coloured layers are enriched in clay and silt. Note that zones of light layers and dark layers define a cycle of about 20 years.

present day. The entire sequence of varves was collected, counted and measured, and a 10,400-year time series for varve thickness (Fig. 1a) was reconstructed⁸. Changes in varve thickness over the past >9,000 years were characterized by constructing an evolutionary spectrum (Fig. 2a) obtained by computing sequential and overlapping power spectra^{11,12} for annual data in 600-year segments of the time series, using a spectral program adapted at Brown University and with the starting time T_0 at calendar year 1927. Increased variance that accompanied the influx of clay and silt, evident in Fig. 1, is accompanied by increased power spectral density from 8,400 to 3,800 varve years (Fig. 2a). The evolutionary spectrum of varve thickness, for the segment between 7,300 and 5,300 years, shows increased spectral density at a period of ~200 years, and also shows an increase at two higher frequencies. The next higher frequency corresponds to a cycle period of ~40–50 years and is expressed for more than 1,500 years, shifting from a lower to a higher frequency over time. A third increase in spectral density occurs near the period of 22 years, over an interval of ~900 years, and is highest at ~6,000 years.

The 9,000-year record of $\Delta^{14}\text{C}$, compiled from tree rings^{1,2}, is a proxy for changes in solar activity because it is linked to the cosmic-ray flux, which in turn is modulated by interplanetary magnetic shocks from CMEs and by other magnetic field changes in the solar wind^{1,2,13}. Deviations are closely linked to the 11- and 22-year solar cycles and also to Maunder-type changes in solar activity¹⁴. I constructed an evolutionary spectrum for $\Delta^{14}\text{C}$ (Fig. 2b), using methods described above, from a decadal time-series derived from tree rings and detrended by a fifth-order polynomial (C. P. Sonett, personal communication).

The association between varve thickness (palaeowinds) and radiocarbon (solar activity) occurs only within a 2,000-year window, between 7,300 and 5,300 years (Fig. 1a). As does varve thickness, the spectrum for ^{14}C within this window displays a three-period structure, with increased spectral density at ~200 years, between 40 and 50 years and between 20 and 25 years (compare *a* and *b* in Fig. 2). Maxima in spectral density for the higher frequencies (periods of 20–25 years and 40–50 years) are not congruent in time as they are for the ~200-year period, but are contained within a broader frequency band. An overlay of the two evolutionary spectra, for periods greater than 50 years in the 7,300–5,300-year window (Fig. 3a), illustrates the congruence of high spectral density at a period of ~200 years. The

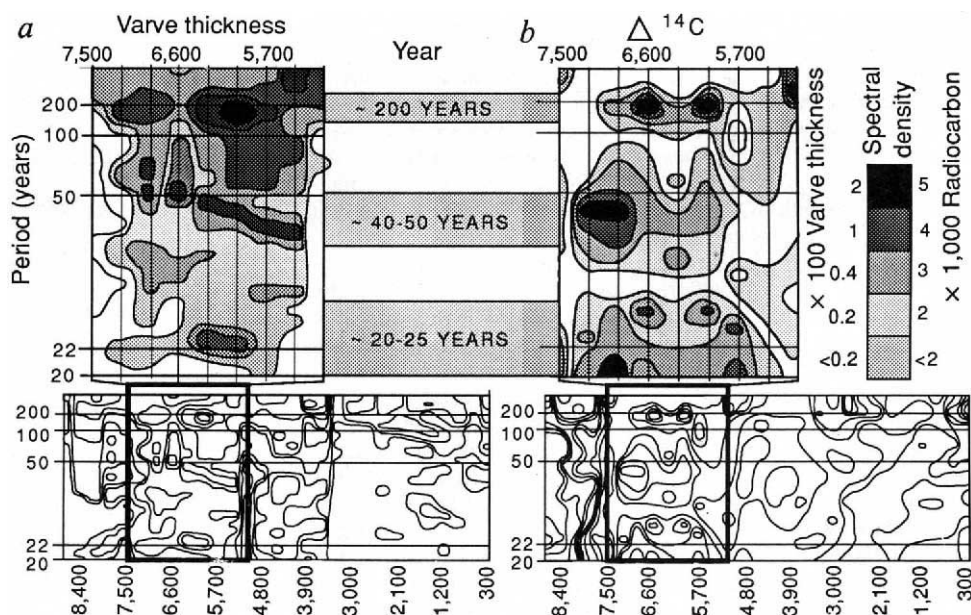
~200-year period persists over the same 2,000-year interval in both spectra and is best developed ~6,000 years ago. This is an interval when the radiocarbon record (>9,000 years long) contains prominent deviations similar to the set of cycles developed during Wolf, Sporer and Maunder episodes of weak solar activity (see Fig. 4 in ref. 1).

I calculated a cross-correlation with radiocarbon for the interval 7,300–5,300 years. The accuracy of the tree-ring chronology within the correlated interval is probably a few decades¹⁵. The accuracy for the varve time series is much lower, probably a few centuries. I initially placed the two records in time by aligning the radiocarbon chronology, based on tree rings, with the Elk Lake chronology obtained from counting varves⁸. Cross-correlation for a segment between 7,300 and 6,300 years has a lag of 50 years and has a coefficient that is significant at the 95% confidence level (Fig. 4b). The long-term trend in varve thickness and radiocarbon (Fig. 4a) is part of a longer, 2,100–2,400-year cycle in ^{14}C (refs 16–18). Removal of the trend by first difference preserved the ~200-year oscillations, and the correlation remained above the 95% confidence level. Cross-correlation for the segment between 6,300 and 5,300 years is below the 95% confidence level, possibly because this segment contains several coring breaks, and some thick varves lose definition and are difficult to enumerate. A significant cross-correlation does not mean that a real temporal correlation has been identified, but visual agreement for principal oscillations, a relatively high coefficient for cross-correlation, and age control afforded by tree rings and varves indicate that the relative positions of the two time series (Fig. 4a) cannot be far from true alignment.

We need to explain why the association is restricted to a 2,000-year window in a continuous record of ~9,000 years. In the case of the time series for varve thickness, only part of the sediment sequence contains sufficient aeolian material to record the effects of palaeowind activity. Before 8,400 years and after 3,800 years, varves contain little aeolian material, and changes in varve thickness are the result of variations in nutrient supply, diatom abundance, water chemistry and lake circulation⁸, factors that cannot be tied directly to wind activity. Even though the association is confined to a 2,000-year window, there is a longer interval of 4,600 years when changes in varve thickness could have recorded such an association.

In the case of the time-series for ^{14}C , changes in the production of radiocarbon have a long-term terrestrial component due to

FIG. 2 Evolutionary spectrum for time series of varve thickness and radiocarbon. *a*, Increased spectral density within the shaded time window in the varve-thickness spectrum occurs at a period of ~200 years, systematically shifts from ~50 to ~40 years, and also occurs at ~22 years; *b*, spectrum for radiocarbon in the window has increased spectral density at ~200 years, with general increases between 40 and 50 years, and 20 and 25 years. Note that increased density at the above three periods is restricted to the indicated 2,000-year time windows.



millennial changes in the Earth's dipole moment, as well as the component that is related to solar activity¹. This relationship with dipole moment is seen in Stuiver's ¹⁴C record from Lake of the Clouds¹⁹ (Fig. 3b), which shows a marked increase in production of radiocarbon in the mid-Holocene, attributed to weakening of the terrestrial magnetic field when the dipole moment reached its lowest value in the Holocene, about 6,000 years ago^{6,7}. A weaker dipole moment allows more lower-energy cosmic rays to enter the atmosphere, producing more radiocarbon. The expression of a three-period spectral structure and Maunder-type deviations in ¹⁴C at that time is to be expected, as the flux of lower-energy cosmic rays is modulated more than that at higher energies by changes related to solar activity in the interplanetary magnetic field²⁰.

The association between CMEs and daily zonal winds³⁻⁵ has been attributed to cosmic-ray flux affecting cloud microphysical processes, and consequently the general atmospheric circulation²¹. These or other atmospheric processes could be effective at longer timescales also, straightforwardly explaining the correlation of palaeowind activity with ¹⁴C variations, including expression of a three-period spectral structure in winds when the dipole moment was weakest. But phenomena other than those directly related to cosmic-ray flux also would be effective at several timescales and may explain or contribute to associations between solar activity and tropospheric winds²².

The association found at Elk Lake is only the second example of a possible connection between climate and the ¹⁴C proxy for solar activity expressed at the Maunder-event timescale, the other example being bristlecone pine²³⁻²⁵. Additional examples will be needed to confirm that the association found at Elk Lake

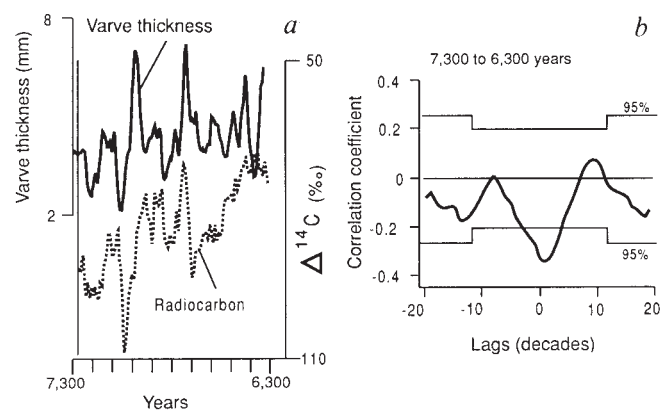


FIG. 4 Plot of changes in varve thickness and secular variation in radiocarbon between 7,300 and 6,300 years ago. *a*, Radiocarbon values are plotted inversely and the radiocarbon time series has been shifted forward 50 years relative to the varve time series to illustrate the best correlation. *b*, Cross-correlation for the shifted time-series segment shown in *a*.

is real. The combination of effects observed at Elk Lake does seem plausible in terms of known geophysical processes and because Maunder-type conditions (quiet Sun) are believed to have a profound effect on solar magnetic flux²⁶ and the interplanetary magnetic field²⁷. Also, the combination of Maunder-type and 22-year oscillations, as seen in the palaeowind record (Fig. 2a), is seen in models that describe long-term changes in the solar wind and the transport of cosmic rays in the heliosphere²⁸. The amplification of cosmic-ray and wind effects that may have occurred when the Earth had a weak magnetic field leaves open the possibility that changes in the geomagnetic field, not directly related to solar activity, may play a role in changing the Earth's climate^{29,30}.

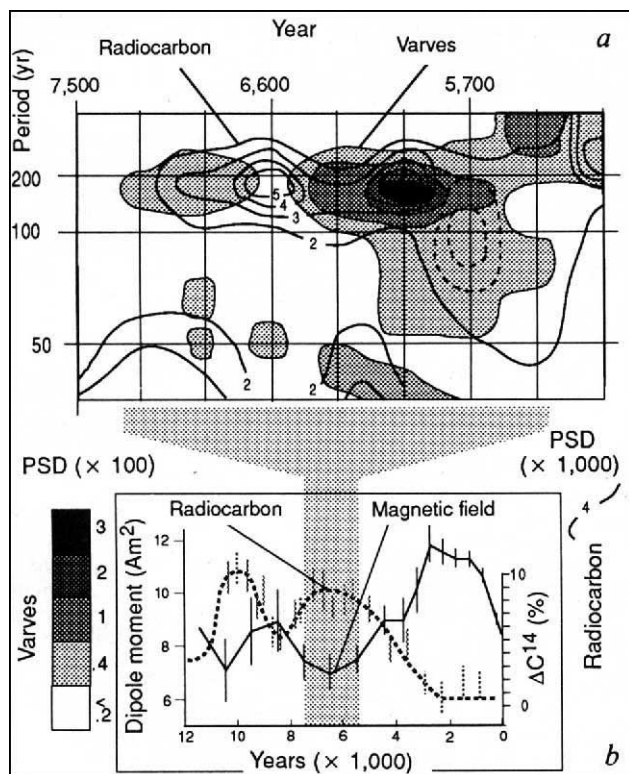


FIG. 3 Comparison of evolutionary spectra for varve thickness and radiocarbon with geomagnetic dipole moment⁶ and production of atmospheric radiocarbon¹⁹. *a*, Spectrum for varve thickness (shaded spectral values), for periods greater than 50 years, is overlain by spectrum for radiocarbon (solid lines). Generally high spectral densities persist in both spectra over a 2,000-year time interval. *b*, Higher spectral densities for varve thickness (palaeowinds) and radiocarbon (cosmic-ray flux) at ~200 years correspond to a time interval with weak dipole moment and increased production of atmospheric radiocarbon.

Received 11 November 1991; accepted 18 May 1992.

- Damon, P. E., Cheng, S. & Linick, T. W. *Radiocarbon* **31**, 697-703 (1989).
- Stuiver, M., Braziunas, T. F., Becker, B. & Kromer, B. *Quat. Res.* **35**, 1-24 (1991).
- McDonald, N. J. & Roberts, W. O. *J. geophys. Res.* **65**, 529-534 (1960).
- Roberts, W. D. & Olson, R. H. *Rev. Geophys. Space Phys.* **11**, 731 (1973).
- Stolov, H. L. & Shapiro, R. *J. geophys. Res.* **79**, 2161-2170 (1974).
- Merrill, R. T. & McElhinny, M. W. *The Earth's Magnetic Field, its History, Origin, and Planetary Perspective* (Academic, London, 1983).
- McElhinny, M. W. & Senanayake, W. E. *J. Geomag. Geoelect.* **34**, 39-51 (1982).
- Dean, W. E., Bradbury, J. P., Anderson, R. Y. & Barnosky, C. J. *Science* **226**, 1191-1194 (1984).
- Grigal, D. F., Severson, R. C. & Goltz, G. E. *Geol. Soc. Am. Bull.* **87**, 1251-1254 (1976).
- Keen, K. L. & Shane, L. C. K. *Geol. Soc. Am. Bull.* **102**, 1646-1657 (1990).
- Anderson, R. Y. & Koopmans, L. H. *J. geophys. Res.* **68**, 877-893 (1963).
- Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications*, 171-208 (Holden-day, San Francisco, 1968).
- Smith, E. J. in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Mathews, M. S.) 175-201 (Univ. of Arizona Press, Tucson, 1991).
- O'Brien, K., De La Zorda Lerner, A., Shea, M. A. & Smart, D. F. in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Mathews, M. S.) 809-831 (Univ. of Arizona Press, Tucson, 1991).
- Stuiver, M., Kromer, B., Becker, B. & Ferguson, C. W. *Radiocarbon* **28**, 969-979 (1986).
- Damon, P. E. in *Secular, Solar, and Geomagnetic Variations in the Last 10,000 Years* (eds Stephenson, F. R. & Wolfendale, A. W.) 267-285 (Kluwer, Dordrecht, 1988).
- Sonett, C. P. & Finney, S. A. *Phil. Trans. R. Soc. A330*, 413-426 (1990).
- Wigley, T. M. L. & Kelly, P. M. *Phil. Trans. R. Soc. A330*, 547-560 (1990).
- Stuiver, M. in *Radiocarbon Variations and Absolute Chronology Proc. 12th Nobel Symp.* (ed. Olsson, I. U.) (Wiley, New York, 1970).
- Lezniak, J. A. & Webber, W. R. *J. geophys. Res.* **76**, 1605-1624 (1971).
- Tinsley, B. A. & Deen, G. W. *J. geophys. Res.* **96**, 22283-22296 (1991).
- Reid, G. C. *J. geophys. Res.* **96**, 2835-2844 (1991).
- Sonett, C. P. & Suess, H. E. *Nature* **307**, 141-143 (1984).
- Suess, H. E. & Linick, T. W. *Phil. Trans. R. Soc. A330*, 403-412 (1990).
- Thomson, D. J. *Phil. Trans. R. Soc. A330*, 601-616 (1990).
- Baliunas, S. in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Mathews, M. S.) 809-831 (Univ. of Arizona Press, Tucson, 1991).
- Suess, S. T. *Planet. Space Sci.* **27**, 1001-1013 (1979).
- Jokipii, J. R. in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Mathews, M. S.) 205-220 (Univ. of Arizona Press, Tucson, 1991).
- Wollin, G., Kukla, G. J., Ericson, D. B., Ryan, W. B. F. & Wollin, J. *Nature* **242**, 34-36 (1973).
- Sternberg, R. S. & Damon, P. E. *Nature* **278**, 36-38 (1979).

ACKNOWLEDGEMENTS. I thank W. E. Dean and J. Platt Bradbury for assistance in compiling Elk Lake varve data, C. P. Sonett for ¹⁴C data, and P. E. Damon, J. O. Fletcher, M. Stuiver and B. A. Tinsley for comments on the manuscript. This work was completed under a NSF grant.

Evidence from Cd/Ca ratios in foraminifera for greater upwelling off California 4,000 years ago

A. van Geen, S. N. Luoma, C. C. Fuller, R. Anima, H. E. Clifton & S. Trumbore*†

US Geological Survey MS 465, 345 Middlefield Road, Menlo Park, California 94025, USA

* Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

† Present address: Department of Geosciences, University of California, Irvine, California 92717, USA

UPWELLING of nutrient-rich Pacific deep water along the North American west coast is ultimately driven by the temperature difference between air masses over land and over the ocean. The intensity of upwelling, and biological production in the region, could therefore be affected by anthropogenic climate change. Examination of the geological record is one way to study the link between climate and upwelling. Because Pacific deep water is enriched in cadmium, dissolved cadmium concentrations in coastal water off central California reflect the intensity of upwelling. By demonstrating that the Cd/Ca ratio in the shell of a benthic foraminifer, *Elphidiella hannah*, is proportional to the Cd concentration in coastal water, we show here that foraminiferal Cd/Ca ratios can be used to detect past changes in mean upwelling intensity. Examination of a sediment core from the mouth of San Francisco Bay reveals that foraminiferal Cd/Ca decreased by about 30% from 4,000 years ago to the present, probably because of a reduction in coastal upwelling. This observation is consistent with predictions of atmospheric general circulation models that northwesterly winds, which drive upwelling, became weaker over this period as summer insolation of the Northern Hemisphere decreased.

Strong winds blow south along the north American west coast during spring and summer because air masses over land become warmer than over the ocean¹. The associated advection of surface water offshore is compensated by upwelling of Pacific deep water along the coasts of California, Oregon and Mexico. The composition of coastal water was strongly affected by upwelling

at Pillar Point and Moss Beach in 1990 and 1991 (Fig. 1). Salinity, phosphate and Cd concentrations rose rapidly as northwesterly winds became stronger in April (Fig. 2). Maxima in May correspond to enrichments of these constituents observed in Pacific deep water 50 km offshore at a depth of ~300 m (ref. 2). The parallel increase of all three tracers also indicates that input of Cd and phosphate by shelf sediment diagenesis was not significant off central California. Salinity, phosphate and Cd returned to their previous values in late summer as upwelling winds became weaker (Fig. 2). The correlation between monthly averaged dissolved Cd concentrations and an upwelling index at 36° N calculated from the monthly averaged atmospheric pressure field for this period (ref. 3 and D. M. Husby, personal communication) was highly significant: the relation $C_{Cd} = 0.25(\pm 0.03) + 1.7(\pm 0.2) \times 10^{-3} \times U$, where C_{Cd} is concentration in nmol kg⁻¹ and U is upwelling in m³ s⁻¹ per 100 m coastline, gives $R^2 = 0.84$, $n = 12$. The mean dissolved Cd concentration during 1990–1991 was 0.42 nmol kg⁻¹, and the mean upwelling index for this period was very close to the long-term average value since compilation started in 1946 (ref. 3).

A relation between Cd/Ca ratios in biogenic calcium carbonate and dissolved Cd concentrations in sea water has been shown previously for deep-ocean benthic foraminifera^{4,5} and scleractinian coral from the Galapagos Islands⁶. *Elphidiella hannah* is a benthic foraminifer that lives in coastal water shallower than 50 m along the north American west coast^{7–9}. The composition of its calcitic shell also reflects Cd concentrations in ambient sea water (Fig. 3). We show this with samples collected from surface sediments at three locations. The first site is Pillar Point (Fig. 1) where the average Cd/Ca ratio of shells collected in rocky pools at low tide is 228 ± 13 nmol per mol ($n = 25$). The distribution coefficient for *E. hannah* is: $K_d = (Cd/Ca)_{shell} / (Cd/Ca)_{sea\ water} = 5.3 \pm 0.3$ (average Cd/Ca of sea water at Pillar Point is 43 nmol per mol). The second site is Richardson Bay, a protected inlet near the mouth of San Francisco Bay (Fig. 1). Here, the mean Cd/Ca ratio for *E. hannah* in surface sediments (362 ± 41 nmol per mol, $n = 6$) reflects an anthropogenic Cd enrichment in the water column throughout the year of ~0.2 nmol kg⁻¹ relative to coastal water, because of anthropogenic inputs within San Francisco Bay (ref. 10, and unpublished data). The third calibration point is less well constrained. A single Cd/Ca determination (132 nmol per mol) is available for *E. hannah* from surface sediment off the Washing-

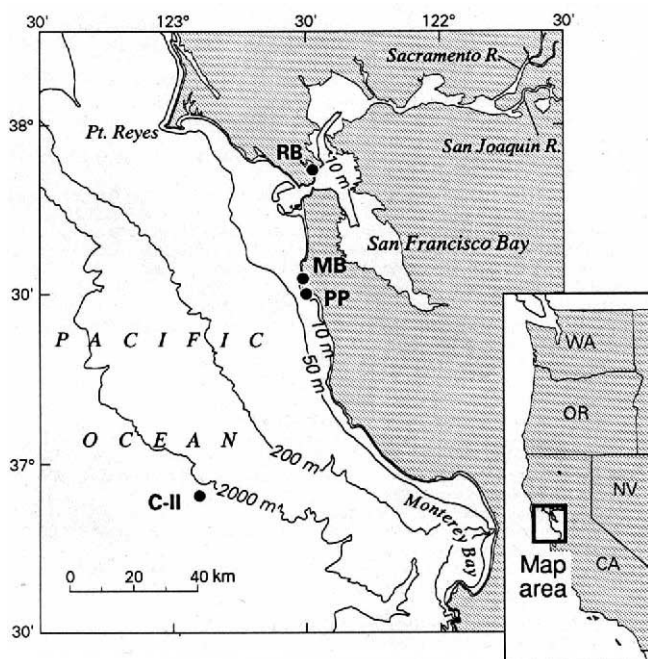


FIG. 1 Location of the sediment section in Richardson Bay (RB) and the near-shore water sampling sites at Moss Beach (MB) and Pillar Point (PP). A vertical profile collected by K. W. Bruland in July 1978 (ref. 2; location C-II in the map) shows that concentrations of upwelling tracers increase roughly linearly from the surface to a depth of 0.5 km within the California Current: 33.3 to 34.2‰ for salinity, 0.6 to 3 $\mu\text{mol kg}^{-1}$ for phosphate and 0.2 to 1.0 nmol kg⁻¹ for dissolved Cd. The composition of near-shore water sampled at the surface in May–June 1990 at PP and MB suggests that it originated from ~300 m depth at C-II.

ton coast where, according to wind patterns, upwelling is considerably weaker than off central California^{1,3}.

Our reconstruction of past upwelling off central California is based on a composite sediment section from Richardson Bay. We collected a 2.5-m-long gravity core in 6 m water depth and retrieved nine 0.5-m core sections nearby by drilling up to 11 m into the sediment (Fig. 4). Radiocarbon ages of five mollusc shell fragments measured by accelerator mass spectrometry (AMS)^{11,12} indicate that the sedimentation rate was $0.26 \pm 0.01 \text{ cm yr}^{-1}$ over the length of the record ($R^2 = 0.989$). High Cd/Ca ratios in foraminifera from the top 0.6 m of the gravity core coincide with the penetration into the sediment of ^{137}Cs produced by atmospheric bomb testing. These elevated values probably reflect an anthropogenic Cd/Ca signal related to discharges into San Francisco Bay over the past century.

The Cd/Ca ratio decreases from 374 ± 16 ($n = 5$) nmol per mol at 11 m depth in the sediment core to a mean pre-industrialization foraminiferal ratio of 278 ± 22 nmol per mol ($n = 75$) between 1 and 2.5 m depth (Fig. 4). We believe that this reflects a change in the composition of coastal water outside San Francisco Bay. It seems that Richardson Bay was not strongly affected by local Cd inputs before it was disturbed by man. Conservative mixing of river water (Cd, $0.1 \times 10^{-9} \text{ mol kg}^{-1}$; Ca, $4 \times$

$10^{-4} \text{ mol kg}^{-1}$) with coastal water under present upwelling conditions (Cd, $0.42 \times 10^{-9} \text{ mol kg}^{-1}$; Ca, $9.8 \times 10^{-3} \text{ mol kg}^{-1}$) would predict a foraminiferal Cd/Ca ratio of $\sim 240 \text{ nmol kg}^{-1}$ for Richardson Bay before anthropogenic disturbance. This value is within the range observed in the gravity core between 1 and 2.5 m depth (Fig. 4). The comparison also shows that differences between the estuarine and the coastal environment (for example, salinity or sediment type) have little effect on the relation between dissolved and foraminiferal Cd/Ca.

A drop in San Francisco Bay salinity linked to higher discharge from the Sacramento and the San Joaquin rivers (Fig. 1) could not explain the elevated Cd/Ca ratio of 4,000-year-old shells. Ratios of $^{87}\text{Sr}/^{86}\text{Sr}$ measured in mollusc shells from our core constrain past salinities in Richardson Bay to a range of 20–30‰ over the past 4,000 years (ref. 13 and B. L. Ingram, personal communication). To increase dissolved Cd/Ca enough to cause the observed change in foraminiferal Cd/Ca at 11 m depth, river water would have to dilute the coastal water to a salinity of 7‰, well below the salinity range tolerated by *E. hannah*. Moreover, pollen data suggest that the effective moisture level in western north America at the time was comparable to or lower than today^{14–16}.

The elevated foraminiferal Cd/Ca ratios in Richardson Bay 4,000 years ago suggest that dissolved Cd concentrations in adjacent coastal water averaged $0.57 \text{ nmol kg}^{-1}$. Recorded fluctuations in yearly averaged upwelling show that this is a physically reasonable value. Coastal upwelling was particularly weak in 1950 and strong in 1955 (ref. 3). From the relation between monthly averaged upwelling and dissolved Cd defined earlier, we would predict mean Cd concentrations of 0.37 and $0.52 \text{ nmol kg}^{-1}$, respectively, for these years. It therefore seems

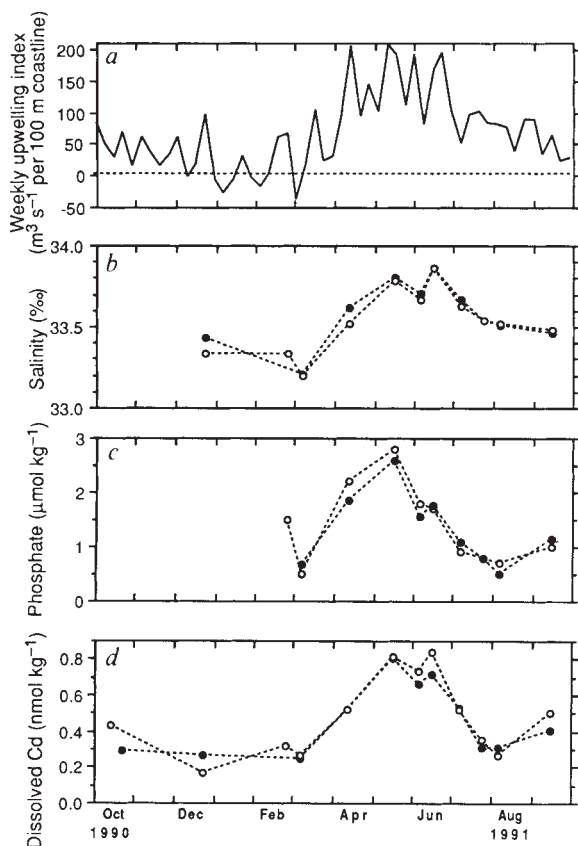


FIG. 2 Time series for weekly upwelling index, salinity, phosphate and dissolved Cd concentrations in near-shore sea water collected between October 1990 and September 1991. Open symbols, Moss Beach; filled symbols, Pillar Point. The upwelling index was calculated at NOAA/Pacific Environmental Fisheries Group (Monterey, California) from the weekly averaged atmospheric pressure distribution at 36°N (ref. 27, and D. M. Husby, personal communication). Note that this measure of upwelling is different from the index discussed in the text, which was based on the monthly averaged atmospheric pressure field³. Coastal water samples were collected from shore and analysed following procedures described in refs 28 and 29. Dissolved Cd, Cu, Ni and Zn concentrations measured at Pillar Point and Moss Beach show that there is little interaction between the water column and shelf sediments for these elements. Mixing with San Francisco Bay water has a negligible effect on phosphate and Cd at Pillar Point and Moss Beach²⁹.

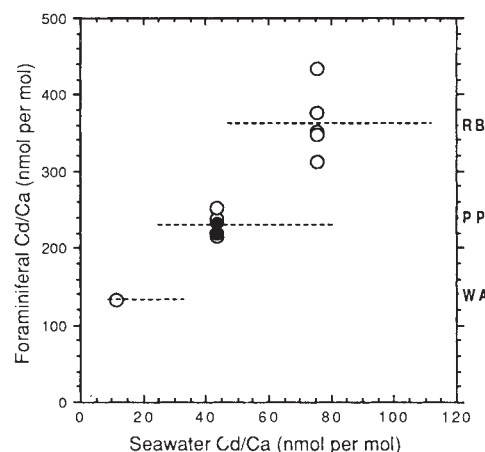


FIG. 3 Calibration of Cd/Ca response in *E. hannah* to ambient dissolved Cd concentrations. Open symbols show each foraminiferal measurement as a function of yearly mean dissolved Cd/Ca at Richardson Bay and Pillar Point. Filled symbols show the composition of Rose Bengal stained shells³⁰ collected in July 1991 and January 1992 at Pillar Point; during the preceding three-month periods the average dissolved Cd concentrations were 0.7 and 0.4 nmol kg^{-1} , respectively. Despite the considerable difference in coastal water composition, the Cd/Ca ratios of stained shells in July and in January are indistinguishable: 223 ± 8 ($n = 7$) and 222 ± 9 ($n = 4$). From this it seems that possible interactions between the life cycle of *E. hannah* and seasonal variations in dissolved Cd in coastal water do not bias the Cd/Ca ratio in the batches of 10–15 shells required for each determination. There was also no difference between the composition of unstained juvenile ($<0.5 \text{ mm}$) and mature ($>0.65 \text{ mm}$) foraminifera shells from Pillar Point. We cleaned and analysed shells following a slightly modified version of the procedure described in ref. 31. The seasonal range in measured dissolved Cd/Ca is shown by a horizontal dashed line at RB and PP. For the Washington coast sample, foraminiferal Cd/Ca is shown as a function of single available dissolved Cd measurement for the region³². A possible underestimate of mean dissolved Cd/Ca for this location is suggested by the estimated seasonal range obtained from the relation in the text and the mean monthly upwelling data for the Washington coast³.

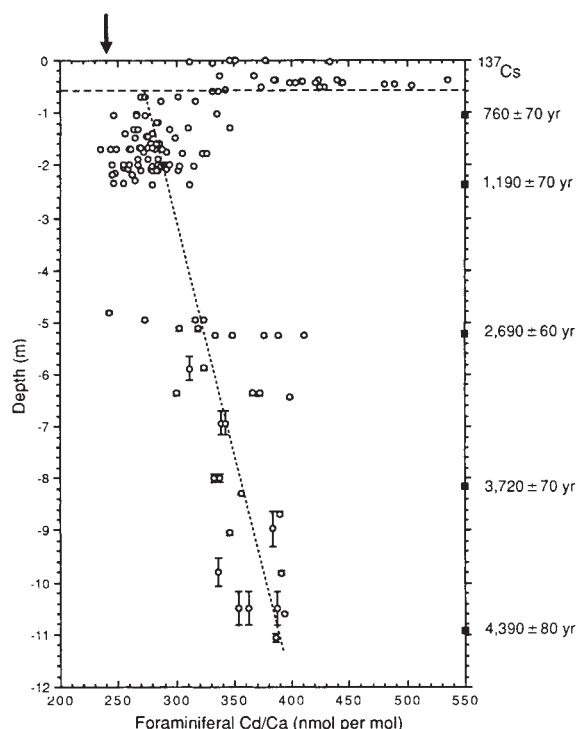


FIG. 4 Composite record of Cd/Ca in *E. hannah* for San Francisco Bay extending to 4,000 years before present. The predicted composition of pre-anthropogenic shells from Richardson Bay is indicated by the vertical arrow. This calculation assumed modern upwelling intensity, conservative mixing and a salinity of 27‰ before river flow control and diversion for agriculture³³ (~6.5‰ lower than the mean salinity of the coastal water endmember). Symbols above the horizontal dashed line are within the zone of penetration of bomb-produced ¹³⁷Cs and reflect anthropogenic discharges into San Francisco Bay over the past century. The depth interval covered by shells for each Cd/Ca determination is smaller than symbol size unless indicated by vertical error bar. The dotted line shows a linear regression of Cd/Ca as a function of depth in the 1 to 11 m interval: $\text{Cd/Ca} = 256(\pm 13) + 12.3(\pm 0.9) \times d$, where Cd/Ca is in nmol per mol, and d is depth in metres, $R^2 = 0.64$, $n = 110$. The provenance and age of mollusc shells, dated by radiocarbon accelerator mass spectrometry, is also indicated. Extrapolation of radiocarbon years to the surface yields an age of $400(\pm 140)$ years, slightly lower than the apparent pre-bomb age of northeast Pacific surface waters³⁴. Radiocarbon ages of 800 ± 90 and $1,070 \pm 70$ years for two mollusc shell fragments from the top interval of the drilled core (4.1 to 4.6 m depth) indicate that sediment from a shallower depth was not effectively blown out of the hole casing before the first drilled core section was taken. For this reason, we have omitted this section from the discussion and the Cd/Ca data (299 ± 16 nmol per mol, $n = 8$) from the figure.

that yearly averaged coastal upwelling 4,000 years ago was consistently as high as the maximum value recorded over the past four decades and has been decreasing since then. Temperature anomalies in coastal water are synchronous along the north American west coast, indicating that such variations in upwelling are not just a local feature¹⁷. Positive temperature anomalies, in particular, are associated with El Niño years¹⁸. We cannot exclude the possibility that a deepening of the Cd gradient offshore² unrelated to coastal upwelling might have contributed to decreasing Cd concentrations in coastal water over the past 4,000 years, but this is not a requirement because salinity (and by inference Cd) profiles 95 km offshore from San Francisco Bay were virtually identical below 300 m depth in 1950 and 1955 (refs 19, 20).

Summer insolation of the Northern Hemisphere has decreased gradually by ~8% from a maximum 9,000 years ago, because of variations in the Earth's orbit around the Sun. Calculations from general circulation models indicate that temperatures over north America decreased by 2–4 °C, and that the atmospheric pressure gradient between land and sea, which causes upwelling winds during summer, was reduced¹⁴. A similar mechanism has been invoked to explain decreasing upwelling in the western Arabian Sea over the past 9,000 years as recorded by the faunal distribution of planktonic foraminifera²¹. Our data suggest that upwelling off western north America also responded to the modest decrease in summer insolation of 3% over the past 4,000 years. A resulting reduction in the biological productivity of surface water and, consequently, a decrease in oxygen consumption by decaying organic matter at depth, may explain why deposition of laminated sediment along the central California continental slope ended ~5,000 years ago²². In the Gulf of California, variations in the oxygen isotopic composition of diatoms²³ and the sediment radiocarbon record²⁴ were also attributed to a reduction in upwelling. Greenhouse warming predicted for the next century, comparable in magnitude to summer warming 9,000 years ago¹⁴, would result in a greater land-to-sea atmospheric pressure gradient off California because the heat capacity of land is lower than that of the ocean²⁵. Although the forcing of greenhouse warming would be year-round, rather than seasonal as was the case 9,000 years ago, the sensitivity of upwelling to changes in insolation in the past

suggests that the intensity of coastal upwelling may increase in the future²⁶. □

Received 12 February; accepted 14 May 1992.

- Huyer, A. *Prog. Oceanogr.* **12**, 259–284 (1983).
- Bruland, K. W. *Earth planet. Sci. Lett.* **47**, 176–198 (1980).
- Bakun, A. NOAA Tech. Rep. NMFS SSRF-693 (Seattle, 1973).
- Hester, K. & Boyle, E. A. *Nature* **298**, 260–261 (1982).
- Boyle, E. A. *Paleoceanography* **3**, 471–489 (1988).
- Shen, G. T., Boyle, E. A. & Lea, D. W. *Nature* **328**, 794–796 (1987).
- Lankford, R. R. & Phleger, F. B. *J. foramin. Res.* **3**, 101–132 (1973).
- Angell, R. W. *J. foramin. Res.* **5**, 85–89 (1975).
- Quintero, P. J. & Gardner, J. V. *J. foramin. Res.* **17**, 132–152 (1987).
- Flegal, A. R., Smith, G. J., Gill, G. A., Sañudo-Wilhelmy, S. & Anderson, L. C. *D. Mar. Chem.* **36**, 329–363 (1991).
- Vogel, J. S., Nelson, D. E. & Southon, J. R. *Radiocarbon* **29**, 323–333 (1987).
- Davis, J. C. et al. *Nucl. Instrum. Methods* **B52**, 269–272 (1990).
- Ingram, B. L. & Sloan, D. *Science* **255**, 68–72 (1992).
- COHMAP Science **241**, 1043–1052 (1988).
- Adam, D. P. & West, G. J. *Science* **219**, 168–170 (1983).
- Owen, K. D., Anderson, R. S., Fall, P. L., O'Rourke, M. K. & Thompson, R. S. *Quat. Res.* **24**, 322–332 (1985).
- Roden, G. I. in *Aspects of Climate Variability in the Pacific and Western Americas* (ed. Peterson, D. H.) 93–111 (American Geophysical Union, Washington DC, 1989).
- Breaker, L. C. in *Aspects of Climate Variability in the Pacific and Western Americas* (ed. Peterson, D. H.) 133–140 (American Geophysical Union, Washington DC, 1989).
- Oceanic Observations of the Pacific, 1950* (eds Rakestraw, N. W., Reid, J. L., Jr., Rudnick, P. & Wooster, W.) (Univ. of California Press, Berkeley, 1960).
- Oceanic Observations of the Pacific, 1955* (eds Reid, J. L., Jr., Arthur, R. S. & Bennett, E. B.) (Univ. of California Press, Berkeley, 1962).
- Prell, W. L., Marvill, R. E. & M. E. Luther *Paleoceanography* **5**, 447–457 (1990).
- Gardner, J. V. & Hemphill-Haley, E. *Geology* **14**, 691–694 (1986).
- Julliet-Leclerc, A. & Schrader, H. *Nature* **329**, 146–148 (1987).
- DeMaster, D. J. & Turekian, K. K. *Paleoceanography* **2**, 249–254 (1987).
- Hansen, J. et al. *J. geophys. Res.* **93**, 9341–9364 (1988).
- Bakun, A. *Science* **247**, 198–201 (1990).
- Bakun, A. NOAA Tech. Rep. NMFS SSRF-671 (Seattle, 1975).
- van Geen, A. & Boyle, E. A. *Analyt. Chem.* **62**, 1705–1709 (1990).
- van Geen, A. & Luoma, S. N. *Estuaries* (in the press).
- Corliss, B. H. & Emerson, S. *Deep Sea Res.* **37**, 381–400 (1990).
- Boyle, E. A. & Keilwin, L. D. *Earth planet. Sci. Lett.* **76**, 135–150 (1985/86).
- Jones, C. J. & Murray, J. W. *Limnol. Oceanogr.* **29**, 711–720 (1984).
- Peterson, D. H. in *Aspects of Climate Variability in the Pacific and Western Americas* (ed. Peterson, D. H.) 419–442 (American Geophysical Union, Washington DC, 1989).
- Southon, J. R., Nelson, D. E. & Vogel, J. S. *Paleoceanography* **5**, 197–206 (1990).

ACKNOWLEDGEMENTS. We thank P. Quintero and M. McGann for help in identifying and picking foraminifera, J. Ingle and D. Sloan for discussing San Francisco Bay foraminifera, D. M. Husby for providing upwelling indices and insights into coastal upwelling along the US west coast, J. Hedges for the Washington coast sediment sample, R. Woolsey, W. O'Neill, B. Noakes and R. Sheldon for drilling in Richardson Bay and the Center of Mass Spectrometry at LLNL for running the ¹⁴C samples. The help of colleagues at USGS was crucial during this work. Comments from M. Delaney, J. Gardner, Y. Rosenthal, F. Schwing and G. Shen improved the manuscript.

Dehydration melting and the generation of continental flood basalts

Kerry Gallagher*† & Chris Hawkesworth*

* Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

† Department of Geological Sciences, University College, Gower Street, London WC1E 6BT, UK

CONTINENTAL flood basalt provinces represent important magmatic events, which may contribute significantly to the generation of new crust. In a typical flood basalt province, large volumes of magma are erupted in a short time: for example, at least 10^6 km^3 in 2–3 Myr in the Paraná–Etendeka province of southeast Brazil and northern Namibia. In many areas flood basalts are associated with mantle plumes, but details of their origin, such as the site of major-element melting, remain unresolved. Recent authors^{1–3} have assumed that partial melting took place at the anhydrous peridotite solidus, and thus concluded that during continental extension, more than 95% of the erupted magmas are generated in the sublithospheric upper mantle. In this situation, the distinctive isotope and trace element geochemistry of the basalts is attributed not to this process, but to the addition of low-degree partial melts scavenged from the overlying lithosphere⁴. Here we present an alternative model in which, in the presence of small amounts of water (~0.4%), the observed quantities of melt may be generated entirely within the mantle lithosphere. As rifting proceeds, however, the basalts acquire an increasingly ‘asthenospheric’ chemistry, as melts from the asthenosphere come to dominate those from the lithosphere.

Some continental flood basalt (CFB) magma types in the Indian Deccan Traps are similar to those erupted recently on the island of Réunion, and so there are close geochemical and spatial links between the Deccan CFB and the Réunion hotspot, and an asthenospheric source is widely accepted. But the Mesozoic Gondwana CFB of the Ferrar (Antarctica), Karoo (South Africa) and Paraná–Etendeka have relatively high SiO_2 contents, distinctive trace element ratios (for example, low Nb/Ba, Nb/La), and enriched Nd, Sr and Pb isotope ratios. As such features are not commonly observed in oceanic basalts and they occur in CFB samples considered not to have been contaminated by continental crust^{5–7}, they seem to reflect distinctive source regions in the subcontinental mantle. Such source regions must be old in order to develop the enriched radiogenic isotope ratios, and so most observers have concluded that they are situated within the continental mantle lithosphere, or that part of the mantle that is isolated from the deeper vigorous convection system. In the lamproite model⁴ most of the CFB magmas are generated in the sublithospheric mantle, but as these melts move upwards, they incorporate lamproitic melts from the lithospheric mantle. Because lamproites are small degree melts and have very high incompatible element contents, they will dominate the trace element and isotope compositions of the resultant CFBs. It has been proposed⁴ that an average of 22% by weight of lamproite melt successfully accounts for the generation of CFBs in Nuanetsi (Karoo Province). It has been argued⁵, however, that mixing models like the lamproite model are not so easily reconciled with geochemical data from large areas of other CFB provinces, such as the Paraná in Brazil and the Ferrar in Antarctica. Objections include the necessity of invoking two asthenosphere components, one depleted and one enriched in trace elements, to explain the observation that magmatism evolves with time from a predominantly lithospheric signature to one which, because of the similarity to oceanic basalts, is inferred to be asthenospheric. Additionally, there is general agreement that no suitable small-degree melts have been identified to

provide a suitable endmember for the low-titanium basalts found in large areas of the Gondwana CFB^{5,8}.

An alternative approach to such mixing models is to consider the effects of small amounts of water on models of partial melting during lithospheric extension. It is well known that the addition of water to peridotite markedly lowers its solidus temperature and promotes the crystallization of hydrous phases such as amphibole and phlogopite^{9–11}. Less well constrained are the ways in which melt compositions vary with P, T and $X_{\text{H}_2\text{O}}$, and how water behaves during fractional melting. Fertile mantle can accommodate up to ~0.4% water in the form of amphibole produced by the hydration of clinopyroxene, with the amount of water then depending on the modal abundance of clinopyroxene^{9,11}. Melts produced from such hydrated peridotite would therefore have compositions controlled by amphibole rather than clinopyroxene and, at least at higher degrees of melting, the partial melts will be silica-saturated¹². In fertile mantle, in which 10–20% clinopyroxene has been

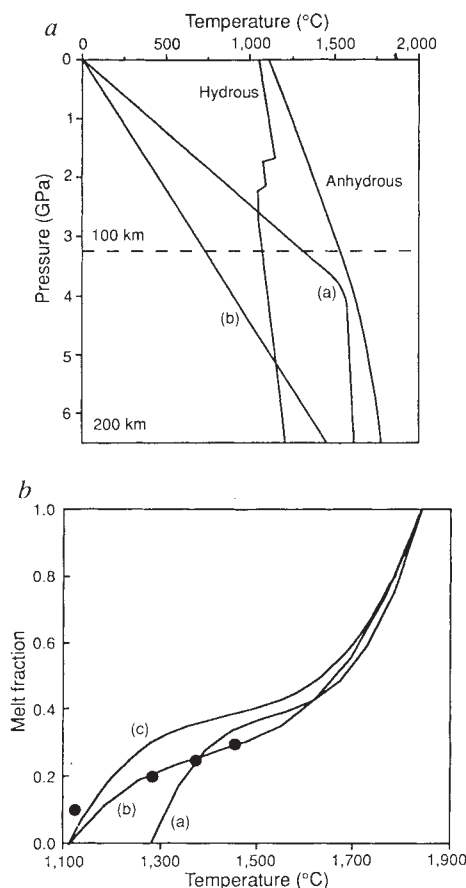


FIG. 1 *a*, The hydrous solidus¹¹ and the anhydrous solidus² used in this work. Also shown are geotherms for a potential temperature of 1,480 °C, and MBL of 100 km (*a*), and 200 km (*b*). The geotherm was calculated using the methodology and basic parameters of ref. 2, with a mantle viscosity of $4 \times 10^{15} \text{ m}^2 \text{ s}^{-1}$. Note that the hydrous solidus is intersected in the lower part of the MBL in both cases. A geotherm with a mantle potential temperature of 1,280 °C also just intersects the wet solidus for both MBL, whereas one with 1,180 °C does not. *b*, Melt fraction, X , at 1.35 GPa (filled circles) for batch melting of hydrous peridotite (derived from Fig. 3a of ref. 11) as a function of temperature. The curves represent functions in the non-dimensional temperature, T' ($T' = [T - 0.5(T_1 + T_2)] / (T_1 - T_2)$), where T , T_s , T_l are the actual, solidus and liquidus temperatures respectively), of the form $X(T') = 0.5 + T' + (T'^2 - 0.25)(a_0 + a_1 T')$. Curve (*a*) is for the anhydrous peridotite solidus² ($a_0 = 0.4256$; $a_1 = 2.988$), curve (*b*) is for the hydrous peridotite solidus ($a_0 = 0.7658$, $a_1 = 2.228$) and curve (*c*) represents the anhydrous peridotite function from curve (*a*) but calculated using the hydrous peridotite solidus. The liquidus of ref. 2 was used in all cases, the implicit assumption being that small amounts of water do not significantly affect the position of the liquidus.

converted to amphibole, considerable volumes of melt could be produced at the relatively low temperatures of the amphibole or hydrated peridotite solidus. But the continental mantle lithosphere is thought to be relatively depleted in major elements^{13,14}, and so it will contain less clinopyroxene or, in the presence of water, amphibole. Melting a depleted peridotite composition again produces silica-saturated melts because the melts are dominated by the orthopyroxene contribution. Moreover, this effect can be further enhanced in the presence of water, because orthopyroxene then melts incongruently¹⁵, producing residual olivine and a silica-saturated melt, and this reaction dominates the composition of the final magma. Thus the combined effect of the presence of water in amphibole, and a depleted major element composition, indicate that dehydration melting of depleted peridotite, as expected in the lithosphere, will tend to produce silica-saturated basaltic melts. The depleted nature of the continental mantle lithosphere will obviously limit the modal abundance of amphibole, but even with 5–10% amphibole, the amount of melt produced could still be 10–15%, with the remaining major element contribution being derived from orthopyroxene and olivine.

We now address the potential of dehydration melting as a mechanism for generating CFB by melting of a hydrated lithosphere. It is assumed that the mantle lithosphere discussed by petrologists and geochemists is effectively equivalent to the mechanical boundary layer (MBL) and so the calculations were designed to evaluate the contribution of melt from this region, relative to that from the asthenosphere below. The MBL overlies a thermal boundary layer which represents the uppermost region of the convecting mantle and in which the geotherm is nearly linear, although the material is in motion. The implication here, then, is that melting above the asthenosphere (that is, within the MBL) provides a source region for the isotope anomalies that indicate stability over timescales of 1–2 billion (10^9) years, and therefore need to be isolated from the convecting system. Although it is possible to invoke long-lived geochemical heterogeneities at deeper levels in the mantle as an explanation for these data¹⁶, this raises the question of why the characteristic chemical signatures only occur in continental areas, and particularly in the Gondwana CFB.

Experimental data for batch melting of hydrous peridotite¹¹ have been used to construct a function relating melt fraction to temperature, and this is shown in Fig. 1a, with the anhydrous peridotite melt function of ref. 2 for comparison. We consider the melt functions to be independent of pressure, which seems to be justified for anhydrous melting experiments^{2,17}. Following from the experimental results, we incorporate the hydrous peridotite solidus of ref. 11, which has ~0.3% water (and ~0.7% CO_2) held in amphibole or phlogopite. This solidus and an anhydrous peridotite solidus² are given in Fig. 1b, together with a geotherm calculated for a potential temperature of 1,480 °C for MBL of 100 and 200 km. The restricted occurrence of CFB (for example, Paraná-Etendeka) compared with the total length of the associated rift margins suggests an association with a mantle plume. A potential temperature of 1,480 °C is equivalent to mantle ~200 °C hotter than normal^{1,2,18}, and may result in even larger temperature excesses relative to cool cratonic lithosphere. In using this steady-state geotherm, we have implicitly assumed that the thermal structure of the MBL has reached an equilibrium state with the abnormally hot plume. In all cases, however, melting occurs only in the lowermost 25% of the MBL, which requires that only the lower part of the MBL needs to be at significantly elevated temperatures. In practice, appropriate timescales for conductive heating in a 150-km MBL are ~20–40 Myr to achieve 30–50% of the total melt generation possible with the steady-state geotherm. The melt extraction and eruption stage, however, can be much more rapid, and our model implies that magma will only be extracted from the lithosphere when it is extended. For the Paraná CFB, intimately linked to the opening of the South Atlantic, the rate of extraction

is then related to the rate of rift propagation, and the expected duration of magmatism would be <5 Myr on the basis of recent plate reconstructions¹⁹. Another factor to consider is that some of the MBL-derived melt may freeze because adiabatic cooling results in the material cutting across lines of constant melt fraction. This freezing effect is poorly constrained, and here we assume that the liquidus temperature of the melt is 600 °C less than the peridotite liquidus and the solidus is the same as the hydrous peridotite solidus. These assumptions are broadly consistent with the solidus and liquidus for water-bearing tholeiite²⁰. There is a maximum on the hydrous solidus at ~1.65 GPa, and for simplicity we monitor the melt temperature at this pressure. For the melt not to freeze at all, the temperature must stay above the melt liquidus and if the temperature is less or equal to the solidus, the melt solidifies completely. If the melt temperature is between the two limits, then we calculate the change in melt fraction due to freezing by assuming a linear variation between the solidus and liquidus.

Our model calculations for melt generation in the presence of a mantle plume and during extension closely follow the approach used in refs 2 and 21, although we incorporate the hydrous solidus in the MBL and allow for freezing of the dehydration melt. The basic results are summarized in Fig. 2 in terms of total melt thickness generated and proportion of this melt derived from the MBL as a function of lithospheric extension factor, β , between 1.0 and 3.0. In the absence of extension ($\beta = 1.0$), the generation of larger quantities of melt with no asthenosphere contribution is favoured by thicker MBL. This is because more material is above the solidus as a result

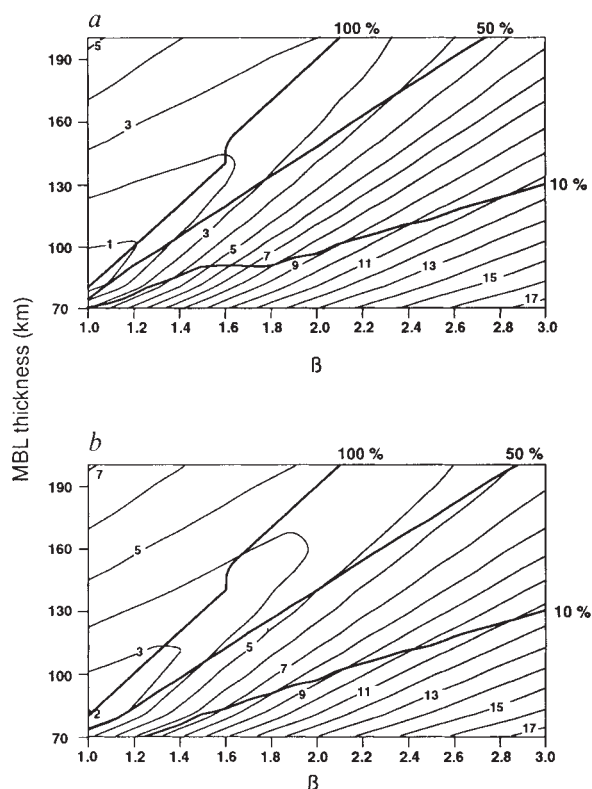


FIG. 2 a, Calculated melt thickness as a function of MBL thickness and extension factor, β , using a geotherm with a potential temperature of 1,480 °C (see Fig. 1a). The contours labelled 100, 50 and 10% represent the percentage contribution of melt derived from the MBL to the total melt thickness and to the left of the 100% contour, all the melt is generated in the MBL. The numerical method used to calculate melting due to adiabatic upwelling during extension closely follows refs 2 and 21, with an entropy change on melting of $250 \text{ J kg}^{-1} \text{ °C}^{-1}$. Freezing of the melt is allowed for, using the approach described in the text. b, As a, but melt freezing is not included.

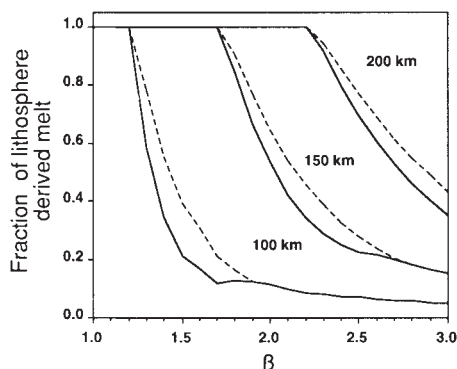


FIG. 3 The fraction of the total melt derived entirely from within the MBL as a function of extension factor, β for 3 values of the MBL thickness. The solid lines include the effects of melt freezing and the dashed lines do not. The actual melt thicknesses can be determined from Fig. 2a and b for the appropriate MBL thickness.

of the impinging hot plume (Fig. 1b). In the current model slightly more than 5 km of melt is predicted for MBL of 200 km, with a maximum of 25% melting at the base of the MBL. The estimated melt thickness increases to over 7 km if no freezing occurs. For more geologically reasonable MBL thicknesses of 150 and 100 km, 3.2 and 1.0 km of melt (or 5.2 and 3.0 km for no freezing) are generated in the absence of extension, the significant feature being that all of the melt is derived from within the MBL. These calculations are too simplistic to estimate the total volume of melt generated, but geometrical arguments imply that for an average of 2 km melt thickness, a volume of $2 \times 10^6 \text{ km}^3$ requires melting to occur over a circular area with a 550-km radius, whereas an average melt thickness of 4 km implies melting over a 400-km radius.

Of greater interest for CFB generation is the situation in which extension of the lithosphere occurs over anomalously hot mantle. In this situation, additional melt is generated through adiabatic decompression and the amount of melt generated across the anhydrous peridotite solidus increases as extension occurs². The corollary for our model is that as the amount of lithospheric extension increases, so too will the asthenosphere contribution. From Fig. 2 we can see that thicker MBL favours a larger proportion of dehydration melt for a given amount of lithospheric extension, although a small amount of MBL-derived melt is predicted in all cases. Extension factors greater than 1.2–1.3 are sufficient to allow asthenosphere melts rapidly to dominate the signature when the MBL is 100 km. But for a value of 200 km, the lithosphere may be thinned by over 200% before the original MBL signature is lost (Fig. 3). Note also that once asthenosphere melting occurs, the contribution from the asthenosphere to the total melt increases more rapidly with extension factor for a thinner MBL. The implication of these results is that, provided the melt can be extracted more rapidly than extension proceeds, there will be time dependence to the composition of the magma erupted at the surface, reflecting the relative contributions from the two source regions in the MBL and the underlying asthenosphere. The results of calculations presented here indicate that for an MBL between 100 and 200 km, 1–5 km of melt may be generated entirely from within the MBL for β values of less than 1.2. As extension progresses (higher β) the proportion of melt from the underlying asthenosphere increases rapidly, as observed for example in both the Paraná and the opening of the South Atlantic⁵, and in the Basin and Range province in the western USA²². □

4. Ellam, R. M. & Cox, K. G. *Earth planet. Sci. Lett.* **105**, 330–342 (1991).
5. Hawkesworth, C. J. et al. *Geol. Soc. Lond. spec. Publ.* (in the press).
6. Mantovani, M. S. M. & Hawkesworth, C. J. *Contrib. miner. Petrol.* **105**, 289–302 (1990).
7. Peate, D. W. thesis, Open Univ. (1989).
8. Hergt, J., Peate, D. W. & Hawkesworth, C. J. *Earth planet. Sci. Lett.* **105**, 134–148 (1991).
9. Green, D. H. *Earth planet. Sci. Lett.* **19**, 36–53 (1973).
10. Mysen, B. O. & Boettcher, A. L. *J. Petrol.* **16**, 549–593 (1975).
11. Olafsson, M. & Eggler, D. H. *Earth planet. Sci. Lett.* **64**, 305–315 (1983).
12. Kushiro, I., *J. geophys. Res.* **95**, 15929–15939 (1990).
13. Boyd, F. R. & Mertzman, S. A. *Geochim. Soc. spec. Publ.* **1**, 13–24 (1987).
14. Hawkesworth, C. J., Kempton, P. D., Rogers, N. W., Ellam, R. M. & van Calsteren, P. W. *Earth planet. Sci. Lett.* **96**, 256–268 (1990).
15. Kushiro, I., Yoder, H. S. & Nishikawa, M. *Bull. geol. Soc. Am.* **79**, 1685–1692 (1968).
16. Gurnis, M. *Geophys. Res. Lett.* **13**, 1474–1477 (1986).
17. Furlong, K. P. & Fountain, D. M. *J. geophys. Res.* **91**, 8285–8294 (1986).
18. Watson, S. & McKenzie, D. *J. Petrol.* **32**, 501–537 (1991).
19. Nürnberg, D. & Müller, R. D. *Tectonophysics* **292**, 27–53 (1991).
20. Green, T. H. in *Andesites: Orogenic Andesites and Related Rocks* (ed. Thorpe, R. S.) 465–488 (Wiley, New York, 1982).
21. McKenzie, D. *J. Petrol.* **25**, 713–765 (1984).
22. Fitton, J. G., James, D. & Leeman, W. P. *J. geophys. Res.* **96**, 13693–13711 (1991).

ACKNOWLEDGEMENTS. We thank M. Bickle, K. Cox, S. Blake and P. Hooper for their comments on the manuscript, and N. Rogers and R. White for discussion.

Marine transgressions and the evolution of Cretaceous dinosaurs

John R. Horner, David J. Varricchio
& Mark B. Goodwin*

Museum of the Rockies, Montana State University, Bozeman, Montana 59717, USA

* Museum of Paleontology, University of California, Berkeley, California 94720, USA

FLUCTUATIONS (transgressions and regressions) of intercontinental sea-ways during the Mesozoic Era affected the habitat area of coastal plains by alternately restricting the space during transgressive phases and expanding the space during regressive phases. In western North America a coastal plain stretched along the eastern front of the young Rocky Mountains during most of the Late Cretaceous. Here we report results of a six-year field study of the sediments and dinosaur remains from this coastal plain in Montana that provide evidence that anagenesis characterized dinosaur evolution at this time, and that the evolutionary pulse coincided with a marine transgression.

In western North America there is a sequence of Upper Cretaceous terrestrial formations with intervening marine strata. Each terrestrial formation contains a unique dinosaur assemblage not found in preceding or succeeding terrestrial units. Morphologically intermediate taxa, which would allow the inference of lineal or near-lineal relationships, between taxa from these terrestrial units, had not been found¹ until recently. To determine whether evolutionary change or migration had occurred to give this unusual pattern^{2,3} the Upper Cretaceous (Campanian) Two Medicine and Judith River Formations of Montana (Fig. 1) were studied for intermediate dinosaur taxa. These formations represent adjacent facies of the Judithian coastal plain of Montana and southern Alberta. The Bearpaw Shale, which overlies the Judith River Formation and much of the Two Medicine Formation, represents a transgressive cycle of the interior sea-way. The 180-m-thick Judith River Formation encompasses roughly 5 million years⁴, whereas the 650-m-thick Two Medicine Formation, based on preliminary $\text{Ar}^{40}/\text{Ar}^{39}$ dates, represents 6–7 million years. Preliminary dates also indicate that the Bearpaw transgression occurred over less than 500,000 years (R. Rogers, J.R.H. and C. Swisher, manuscript in preparation).

Dinosaur remains were collected from upper sediments of the Two Medicine Formation that were deposited during the final stages of the Bearpaw transgression. Several dinosaur taxa found can only be defined by combinations of synapomorphic features compared to primitive taxa, and plesiomorphic features as compared to derived taxa. Because these taxa have no

Received 14 November 1991; accepted 11 May 1992.

1. White, R. S. & McKenzie, D. P. *J. geophys. Res.* **94**, 7685–77302 (1989).
2. McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625–679 (1988).
3. Arndt, N. T. & Christensen, U. *J. geophys. Res.* (in the press).

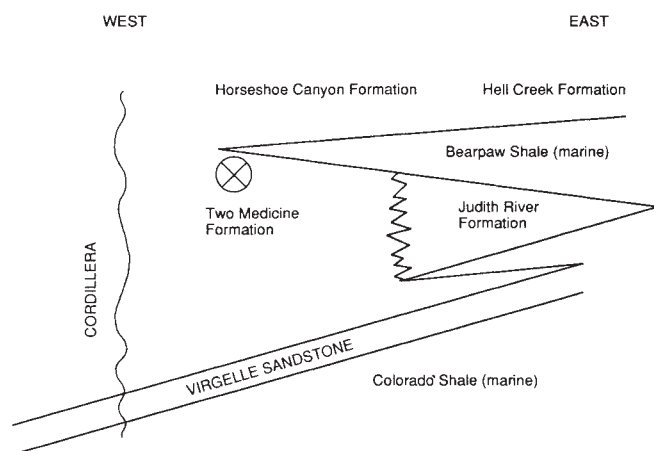


FIG. 1 Stylized cross section of Upper Cretaceous (Campanian–Maastrichtian) strata in north central Montana and southern Alberta. X is the approximate stratigraphic position of the 80-m sections from which the described intermediate taxa were collected.

autapomorphies (unique derived characters) they are interpreted as intermediate or transitional taxa. The transitional ceratopsid and lambeosaurid taxa discussed below are known from numerous individuals collected from nests, monospecific bonebeds, or both. Individual variation is therefore known for these taxa. The tyrannosaurid and pachycephalosaurid taxa were

found as isolated individuals. Each of these dinosaur taxa (to be described elsewhere) seem to be evolutionary intermediates between taxa from the older Judith River Formation and the younger Horseshoe Canyon or Hell Creek Formations.

Anagenesis, rarely documented in terrestrial vertebrates and almost unknown in Mesozoic vertebrates, seems to be clearly evident in these dinosaurs. Three new ceratopsids (Fig. 2a) seem to represent three stages in the evolution from a taxon related to *Styracosaurus* to *Pachyrhinosaurus*. *Styracosaurus*, from the Judith River Formation of Alberta, has a number of autapomorphies, most diagnostic of which is the possession of several parietal spikes on either side of its frill. Transitional taxon A* (represented by six specimens) has cranial and postcranial features that are identical to *Styracosaurus*, except that it possesses only one pair of parietal spikes. Transitional taxon B* (20 specimens), also bearing a frill with one pair of parietal spikes, has a nasal horn that curves anteriorly over the premaxillary process of the nasal. Transitional taxon C* (four specimens), with one pair of parietal spikes, has a nasal horn that extends anteriorly over and attaches to the premaxillary process of the nasal. The dorsal surface of this horn is slightly raised and extremely rugose. The surface of bone over the orbits bears coarse ridges. *Pachyrhinosaurus*, from the Horseshoe Canyon Formation, has a pair of parietal spikes, additional parietal ornaments and a nasal boss and orbital bosses with coarse ridges. The three transitional taxa are found in stratigraphic sequence (Fig. 2) and do not overlap in time with one another. There is no evidence of cladogenesis, and therefore nothing to rule out the possibility of anagenesis^{5,6}.

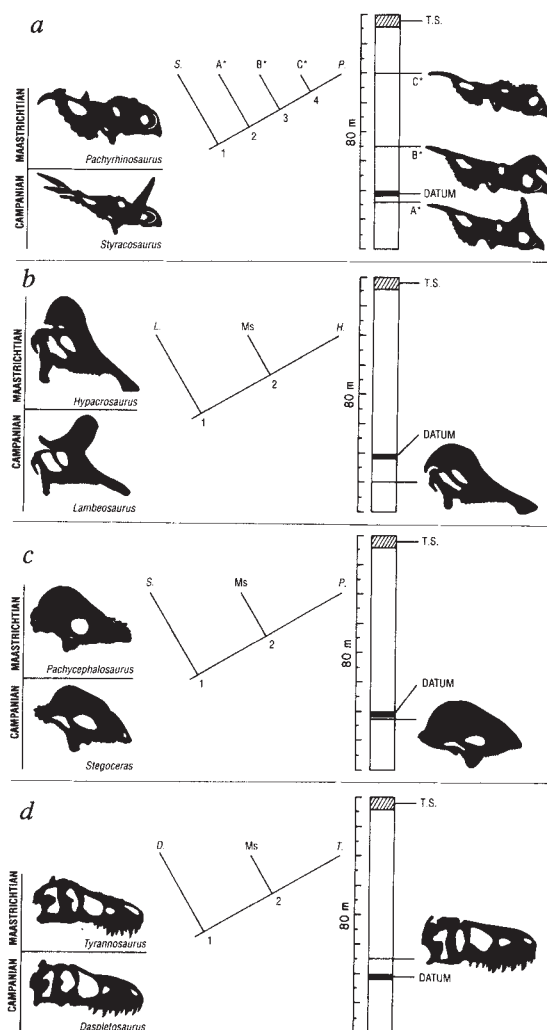


FIG. 2 Cladograms and stratigraphic columns showing relationships and stratigraphic positions of intermediate taxa. Drawings in boxes on left represent taxa proposed to be linked by the intermediate taxa. Stratigraphic columns on right represent the uppermost 80 m of the Two Medicine Formation. TS is the transitional sediments of the Bearpaw Shale. The datum is a bentonite, with an Ar^{40}/Ar^{39} date of 74.3 Myr (R. Rogers, J.R.H. and C. Swisher, manuscript in preparation). Intermediate taxa are to the right of the stratigraphic columns. Representative characters for the numbered nodes on the cladograms are as follows: a, 1, Multiple parietal frill spikes; 2, one pair of parietal frill spikes; 3, anteriorly projecting nasal horn; 4, rugose nasal boss on dorsal surface of nasal horn and rugose orbital bosses. b, 1, Large rounded narial crest with nasal contributing to at least half of lateral surface of crest; 2, broad crest formed half by the nasal and having little or no nasal vacuity, and tall dorsal neural spines. c, 1, Doming of the frontoparietal; 2, pronounced doming of the skull incorporating lateral margins of the cranium into dome with corresponding loss of parietal shelf. d, 1, Long postorbital/frontal suture, inflated ventral jugal processes of ectopterygoid, and anterolateral edge of antorbital ramus of lacrimal not continued across lateral surface of jugal; 2, first antorbital fenestra higher than long, enlarged, basisphenoid ventral foramina, and articular surface of occipital condyle extends far antero-ventrally. Ms, Metaspecies: intermediate taxon.

A new lambeosaurid (represented by more than 50 specimens; Fig. 2b) has cranial characters intermediate between a taxon closely related to *Lambeosaurus*, from the Judith River Formation, and *Hypacrosaurus*, from the Horseshoe Canyon Formation. The most significant character of *Lambeosaurus* is the greatly expanded premaxillae that form most of the narial crest. Autapomorphies of *Hypacrosaurus* include laterally expanded nasals that comprise half of the narial crest, a very restricted narial opening, and dorsal vertebrae with extremely reduced centra and long neural spines. The transitional lambeosaurid has a laterally expanded narial crest, the majority of which is composed of the expanded premaxillae, reduced narial openings, and reduced dorsal centra with long neural spines.

A transitional pachycephalosaurid (represented by three specimens; Fig. 2c) has a frontoparietal dome that is intermediate between that of *Stegoceras* from the Judith River Formation and *Pachycephalosaurus* from the Hell Creek Formation. The intermediate taxon has small, bulbous olfactory lobes immediately anterior to the cerebrum as in *Stegoceras*, plus pronounced frontoparietal doming with the lateral margin of the skull incorporated into the dome, as is characteristic of *Pachycephalosaurus*.

An intermediate tyrannosaurid (represented by three specimens; Fig. 2d) has cranial characters that suggest an evolutionary position between *Daspletosaurus* from the Judith River Formation and *Tyrannosaurus* from the Hell Creek Formation. The intermediate taxa and *Daspletosaurus* have a lacrimal horn that is triangular in lateral aspect, with the apex behind and above the antorbital ramus. Also, the lacrimal horn exceeds the postorbital horn in size. The intermediate taxa and *Tyrannosaurus* share a first antorbital fenestra that is higher than long, enlarged basisphenoid ventral foramina, and an articular surface of the occipital condyle that extends down and ventrally onto the neck of the condyle. The morphologic differences among the above mentioned taxa are largely those that have been argued to be related to sexual display⁷. If so, it could imply that sexual selection was a principal engine of morphologic change in these dinosaurian taxa.

The Judith River Formation was deposited over a 5 million year period, yet throughout these strata there are no identified intermediate dinosaur taxa, and therefore no indication of their evolution during this time span. This also seems to have been the case among the Judithian mammals⁴. There seems to be evolutionary stasis of most if not all dinosaur and mammal taxa from this formation. But in a time span of less than 500,000 years, represented by the top 80 m of the Two Medicine Formation, several dinosaur taxa appear to be intermediates, suggesting that evolutionary transitions coincided with transgression.

Much of the anagenetic change that we observe may have coincided with environmental stress. High sea-way transgressions caused habitat bottlenecks that apparently stimulated evolutionary change in independent lineages of dinosaurs, the evidence for which is shown by changes in sexual ornamentations. □

Quantification of role reversal in relative parental investment in a bush cricket

L. W. Simmons

Department of Zoology, The University of Western Australia, Nedlands, Western Australia 6009, Australia

SEXUAL differences in courtship roles are thought to depend on the ratio of sexually available females to males¹⁻³. This 'operational sex ratio'³ should be biased towards the sex investing least in reproduction because they will have the higher potential reproductive rate⁴. Thus relative parental investment^{1,2} is thought to underlie observed courtship roles. Typically, males have the lower investment and compete for choosy females^{2,3,5}. But role reversal can occur when males invest parentally^{2,3}. Recent evidence supports the contentions that role reversal reflects shifts in the potential reproductive rates of males and females⁴ and shifts in the operational sex ratio⁶. Cases of male investment exceeding female investment in role-reversed species have not been found. In some insects, males invest in reproduction by providing the female with nutrients used for zygote production^{7,8}. Here I report the measurement of male and female investment in reproduction for a nutrient-provisioning bush cricket (Orthoptera: Tettigoniidae) and demonstrate that reversals in relative parental investment underlie courtship role reversals seen in this species^{6,9}.

During mating, male bush crickets transfer a complex spermatophore consisting of a sperm-carrying ampulla and a proteinaceous spermatophylax that is consumed by the female¹⁰. Spermatophylax nutrients increase the number and survival prospects of offspring¹¹⁻¹³. Thus, both males and females invest parentally in eggs by nutrient provisioning. As the availability of food in the environment decreases, there is a decrease in the number of males able to obtain enough nutrients to donate to females^{6,9,14}. Concomitant with decreased male availability is

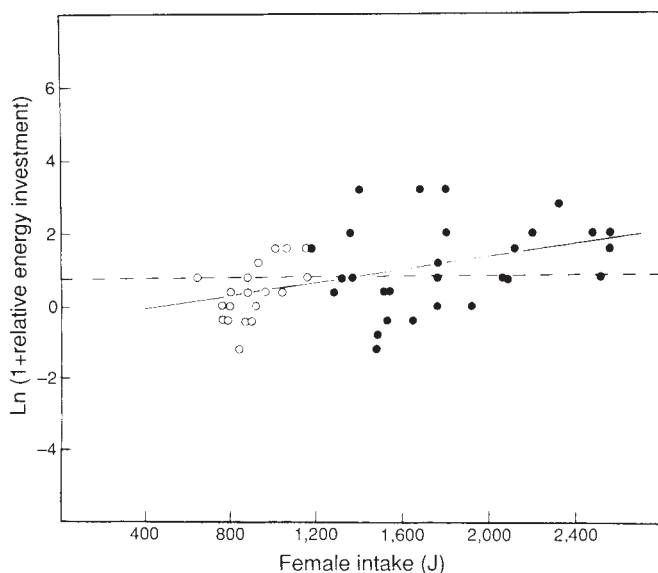


FIG. 1 The relationship between relative energy investment (J per egg for males; J per egg for females) and energy intake of females over a single reproductive episode. Ratios were log transformed for analysis. The dashed line depicts the line of equality in energy investment. ○, Females allowed limited access to pollen; ●, females fed freely. Points below the line correspond to a greater male energy investment, those above to a greater female energy investment (regression analysis $F=10.58$, $d.f.=1, 45$, $P=0.002$, $r^2=0.172$; the line is described by the equation $Y=8.56E-4 \ln(1+X)-0.38$).

Received 8 March; accepted 18 May 1992.

1. Weishampel, D. B. & Horner, J. R. in *4th Symp. Mesozoic Terrestrial Ecosystems* (eds Currie, P. M. & Koster, E. H.) 224-229 (3rd Occ. Paper, Tyrrell Mus. Palaeont., 1987).
2. Bakker, R. T. in *Patterns of Evolution* (ed. Hallam, A.) 431-478 (Elsevier, Amsterdam, 1977).
3. Horner, J. R. in *Montana Geological Society 1984 Field Conference Guidebook* (eds McBane, J. D. & Garrison, P. B.) 299-303 (Montana Geological Society, Billings, 1984).
4. Goodwin, M. B. & Deino, A. L. *Can. J. Earth Sci.* **26**, 1384-1391 (1989).
5. Gingerich, P. D. *Paleobiology* **11**, 27-41 (1985).
6. Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T. J. M.) 82-115 (Freeman, Cooper, San Francisco, 1972).
7. Coombs, W. P. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 32-42 (Univ. California Press, Los Angeles, 1990).

ACKNOWLEDGEMENTS. We thank P. Currie, P. Gingerich and K. Padian for their critical reviews and discussions, R. Rogers, S. Sampson and D. Weishampel for discussions concerning their related data; C. Ancell, B. Harmon, P. Leiggi and B. Taylor for specimen preparation; K. Ellingsen for illustrations; and the Blackfeet Nation for access to their tribal lands. This research was supported by NSF grants to J.R.H.

an increase in female mating frequency as they seek additional courtship meals^{6,9,14}. Thus the operational sex ratio³ becomes skewed towards females who compete for available males⁶. Competition among females may arise because of the scarcity of nuptial males and need not imply that male investment in reproduction exceeds female¹⁵. Nevertheless, the fact that females incur increased time and energy costs associated with remating suggests that the male's investment becomes more important for reproduction^{6,9,14}. I measured the energetic invest-

TABLE 1 Energy budget for a single reproductive event

	Diet	Males (J)	Females (J)
Intake	Unlimited	503.1 ± 27.1 ^a	1,802.8 ± 80.0 ^b
	Limited	454.9 ± 25.9 ^a	902.1 ± 31.2 ^c
Faeces	Unlimited	136.0 ± 8.6 ^d	556.6 ± 26.4 ^e
	Limited	102.7 ± 10.0 ^d	242.4 ± 12.7 ^f
Reproduction	Unlimited	59.5 ± 5.0 ^g	*199.3 ± 28.0 ^h
	Limited	53.6 ± 4.5 ^g	78.5 ± 14.4 ^g
Change	Unlimited	74.6 ± 8.3 ⁱ	251.6 ± 16.2 ^j
	Limited	49.8 ± 10.0 ^k	19.5 ± 19.7 ^k
Metabolized	Unlimited	232.9 ± 20.0 ^l	795.3 ± 58.1 ^m
	Limited	248.7 ± 14.3 ^l	561.7 ± 18.5 ⁿ
MANOVA	Wilk's lambda	F (5, 86)	P
Diet		0.394	26.43
Sex		0.194	71.54
Diet × Sex		0.481	18.54

The species is currently undescribed but catalogued as Gen. nov. 22, sp. 1 in the Australian National Insect Collection²⁴. Animals were collected as penultimate instars from spring flowers in King's Park, Western Australia and reared to adulthood in population cages supplied with an assortment of spring flowers for food (zaprochilines feed exclusively on pollen and nectar²⁵). Each day, newly eclosed animals were removed, weighed, and placed into individual 250 ml jars upturned onto clean petri dishes in order to collect faeces. Animals were assigned at random to the two diet treatments. A dish of 'bee pollen'^{6,9} was weighed and placed into each jar. For the limited diet the dish was removed after 24 h and reweighed to determine the weight of pollen consumed. Animals were thus fed on every third day. For the unlimited diet, food dishes were removed and reweighed at the end of the experimental period. These dietary levels induced reproductive levels in females known to result in courtship role reversal (when dissected at the termination of the experiment, freely fed females had a mean ± s.e. of 14.8 ± 1.28 developing eggs in their ovaries compared to 4.21 ± 0.82 for females on the limited diet)⁶. After the 10 day pre-mating period¹⁶, 28 unlimited diet females and 19 limited diet females were provided with a male of their own dietary regime and allowed to mate. Male weight was taken before and after copulation to estimate spermatophore weight. Males were frozen for use in energy measurements and their faeces weighed. Females were provided with a pot of dry sand in which to oviposit and left for 20 days, the maximum duration of the zaprochiline's refractory period during which oviposition occurs¹⁶. Females were then frozen and the collected faeces and eggs weighed. All weights were converted to dry weights using predetermined wet weight/dry weight relationships. The energy contents of dry samples of pollen, faeces, male bodies, female bodies, eggs and spermatophores were obtained using a Phillipson microbomb²⁶ calibrated with benzoic acid. Weights were converted to their energetic equivalents by multiplying by the appropriate energetic value. For females, energy intake included pollen as well as energy obtained by spermatophylax consumption. Energy change was calculated from the difference between body weight at eclosion and at the end of the reproductive event (after spermatophore donation for males or oviposition for females). Energy metabolized was calculated as the energy intake minus that accounted for by faeces, reproduction and energy change. Within energetic parameters, means with the same letter do not differ significantly at 5% level by Tukey's multiple comparisons.

* Radiolabel experiments show that females on a limited diet would incorporate 70% of male material into eggs over a period of 20 days compared to 88% for females fed freely (L.W.S. and D. T. Gwynne, manuscript in preparation). Thus each female's investment should be reduced by that proportion of their mate's investment; unlimited- and limited-diet females therefore invested 146.9 ± 28.3 J and 40.9 ± 15.8 J of their own energy in reproduction.

ment in reproduction by males and females to test the prediction that male investment exceeds female investment during courtship role reversal.

I calculated energy budgets for male and female zaprochiline bush crickets^{6,9} maintained under two dietary regimes over the course of a single reproductive event (Table 1). This includes a pre-reproductive period after adult eclosion. For females there is also a period after mating, the refractory period^{16,17}, when eggs are matured and laid.

Energy budget parameters differed between sexes and diets (Table 1). Females were expected to have a greater intake and faecal output and to metabolize more energy because their reproductive event lasted longer than that of males. Females also stored more energy for future reproduction (reflected by the energy change) and invested more in current reproduction than did males. But females allowed limited access to pollen had a lower energy intake, and therefore lower energy loss from defecation. Furthermore, their current reproductive investment did not differ from that of males and their future reproductive investment tended to be lower than that of males and females fed freely (lower energy change and number of developing eggs). Diet did not affect male energetic parameters, explaining the sex by diet interactions (Table 1). Males provided unlimited pollen fed sufficiently to invest in a single reproductive event after which they spent their time calling to attract females.

Relative energy investment in offspring was calculated as the ratio of energy spent in reproduction per egg by females to that

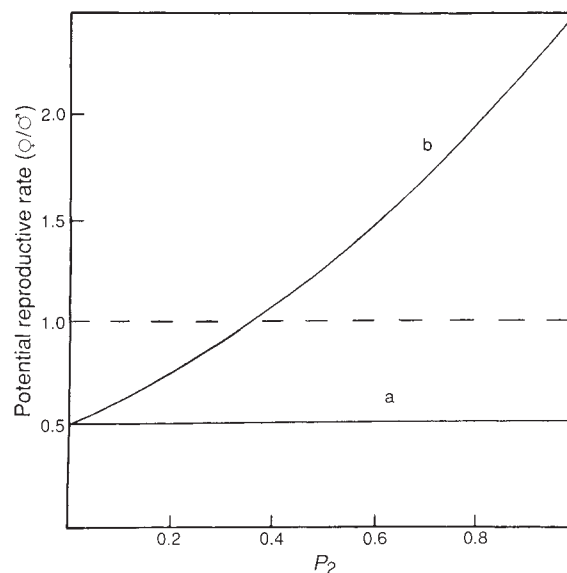


FIG. 2 A model of the effect of varying levels of sperm competition on the maximum potential reproductive rates of females relative to males fed freely (a) and nutrient limited (b). To estimate the maximum potential reproductive rate for a male, the model assumes (1) that each female is a virgin when the male mates and (2) that after each additional mating by the female, the initial male obtains the same proportion of those eggs not fertilized by the last male as he did of the total clutch before the last male copulated. Relative reproductive rate, $R = et / mt(er + \sum_i (1 - P_2)^i er)$, where e = egg production rate of females (measured as 0.38 ± 0.05 eggs per day and 0.18 ± 0.03 eggs per day for the unlimited and limited diets provided in this study, respectively); m = mating rate of males (for heuristic purposes m is assumed equal to 0.1 for both diets because all males were able to mate after 10 days feeding. In reality males fed freely may well be able to mate more frequently^{13,14}); t = period of reproduction (20 days in this study); r = remating interval of females (20 days for well fed females versus 4 days for food-limited females¹⁶); P_2 = proportion of eggs fertilized by the last male; and i = number of matings in time t by the female. Because $r = t$ for females fed freely, $i = 1$, sperm competition is irrelevant and the equation simplifies to $R = et / mter$. Values < 1 indicate a greater potential reproductive rate for males, values > 1 indicate a greater potential reproductive rate for females.

by males. Diet affected relative energy investment ($t = 2.22$, d.f. 45, $P = 0.03$, ratios were log transformed for analysis). For the unrestricted diet, females invested more energy per egg than did males (mean $\pm$ s.e. ratio 5.01 ± 1.48) whereas for the limited diet the mean relative energy investment did not differ significantly from 1 (1.21 ± 0.45 ; $t = 0.47$, d.f. 18, n.s.) indicating that, on average, males and females invested equally in eggs. Inevitably there was variation in the levels of female nutrition within diet treatments. Thus a more accurate analysis of the effect of diet is to plot relative energy investment against the actual energy intake by females (Fig. 1). Female energy investment relative to male increased with energy intake. Furthermore, the regression line in Fig. 1 shows that there was a role reversal in relative energy investment with males investing relatively more per egg when females had an intake of less than 1,242.6 J (equivalent to 59.5 mg pollen over the reproductive event or 2.0 mg pollen per day). This reversal in relative energy investment was a consequence of reduced female investment under nutrient limitation and not increased male investment (Table 1).

These data support the argument that male parental investment increases in value when environmental resources are scarce^{6,9}. Furthermore they provide empirical support for Williams' theory¹ that sexual differences in courtship roles reflect the relative parental investment of the sexes. Trivers² defined parental investment in terms of decrement to the parent's ability to invest in future offspring. The male's investment limits his potential to invest in future offspring because of the time required to recoup spent resources¹³. How current investment affects future reproduction by females is unknown.

Relative parental investment is thought to constrain the potential reproductive rates of the sexes, which in turn determines the operational sex ratio^{4,18}. Although female reproductive rate is set by the rate of egg production, male reproductive rate will be a function of his mating frequency and sperm competition because paternity expectation in insects can decrease when females remate¹⁹. Hungry females will remate sooner than well fed females¹⁶. The outcome of sperm competition in zaprochilines is unknown, but the potential reproductive rates of females to males under different sperm priority patterns can be modelled (Fig. 2). When animals are fed freely, the potential reproductive rate of males will exceed that of females because females rarely remate and a male can provision several females during the reproductive period. When nutrients are limited, potential male reproductive rate quickly falls below that of females as levels of sperm priority increase and males lose fertilizations to the females' other mates. With one possible exception²⁰ those bush crickets studies so far show moderate-to-high second-male sperm precedence²¹⁻²³ such that reversal in relative reproductive rates of the sexes is likely to accompany the reversal in relative parental investment. Together with the female bias in the operational sex ratio⁶, these changes are manifest behaviourally as females cease to be discriminative and compete for choosy males^{6,9}. □

19. Parker, G. A. *Biol. Rev.* **45**, 525-567 (1970).
20. Gwynne, D. T. *Behav. Ecol. Sociobiol.* **23**, 373-377 (1988).
21. Wedell, N. *Evolution* **45**, 1975-1978 (1991).
22. von Helversen, D. & von Helversen, O. *Behav. Ecol. Sociobiol.* **28**, 391-396 (1991).
23. Achmann, R., Heller, K.-G. & Epplen, J. T. *Molec. Ecol.* (in the press).
24. Rentz, D. C. F. *A Monograph of the Tettigoniidae of Australia Vol. 2: The Phamodinae, Zaprochilinae and Saginae* (Crawford House, Bathurst, Australia, in the press).
25. Rentz, D. C. F. & Clyne, D. J. *Aust. Ent. Soc.* **22**, 155-160 (1983).
26. Phillipson, J. *Oikos* **15**, 130-139 (1964).

ACKNOWLEDGEMENTS. I thank P. Withers for use of equipment and for discussion, D. Gwynne, D. Roberts and R. Black for comments on earlier drafts, and L. Partridge whose discussions both prompted these experiments and greatly improved the manuscript. This research was funded by a Queen Elizabeth II fellowship from the Australian Research Council.

Voltage-dependent phosphorylation may recruit Ca^{2+} current facilitation in chromaffin cells

Cristina R. Artalejo*†, Sandra Rossie‡, Robert L. Perlman*§ & Aaron P. Fox*||

* Department of Pharmacological and Physiological Sciences,

§ Department of Pediatrics and the Joseph P. Kennedy Jr Mental Retardation Research Center, The University of Chicago, 947 East 58th Street, Chicago, Illinois 60637, USA

† Departamento de Farmacología, Facultad de Medicina, Universidad Autónoma de Madrid, Madrid 28029, Spain

‡ Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907, USA

BOVINE chromaffin cells have two components of whole-cell Ca^{2+} current: 'standard' Ca^{2+} currents that are activated by brief depolarizations, and 'facilitation' Ca^{2+} currents, which are normally quiescent but can be activated by large pre-depolarizations or by repetitive depolarizations to physiological potentials¹⁻⁵. The activation of protein kinase A can also stimulate Ca^{2+} current facilitation, indicating that phosphorylation can play a part in facilitation⁶. Here we investigate the role of protein phosphorylation in the recruitment of facilitation Ca^{2+} currents by pre-pulses or repetitive depolarizations. We find that recruitment of facilitation by depolarization is a rapid first-order process which is suppressed by inhibitors of protein phosphorylation or by injection of phosphatase 2A into cells. Recruitment of facilitation Ca^{2+} current by voltage is normally reversible but phosphatase inhibitors render it irreversible. Our results indicate that recruitment of these Ca^{2+} currents by pre-pulses or repetitive depolarizations involves voltage-dependent phosphorylation of the facilitation Ca^{2+} channel or a closely associated regulatory protein. Voltage-dependent phosphorylation may therefore be a mechanism by which membrane potential can modulate ion channel activity.

Figure 1a shows the time course for the recruitment of facilitation Ca^{2+} current by pre-pulses. Pre-pulses as brief as 1 ms produced a measurable increase in Ca^{2+} current. The time course for the recruitment of facilitation was well fitted by a single exponential with $\tau = 26$ ms.

The experiment in Fig. 1a was carried out with GTP in the patch pipette solution. To determine whether the recruitment of Ca^{2+} current facilitation by pre-pulses was dependent on activation of G proteins, GTP was omitted from the patch pipette solution. Omission of GTP from the patch pipette blocked the recruitment of facilitation by a D_1 dopamine agonist (SKF 38393), which recruits facilitation through an adenylyl cyclase/protein kinase A cascade⁶, indicating that GTP levels in the cell were too low to support the activation of G proteins (Fig. 1b). Depletion of GTP did not, however, impair the

Received 23 January; accepted 19 May 1992.

1. Williams, G. C. *Adaptation and Natural Selection* (Princeton Univ. Press, 1966).
2. Trivers, R. L. in *Sexual Selection and the Descent of Man 1871-1971* (ed. Campbell, B.) 136-179 (Aldine, Chicago, 1972).
3. Emlen, S. T. & Oring, L. W. *Science* **197**, 215-222 (1977).
4. Clutton-Brock, T. H. & Vincent, A. C. J. *Nature* **351**, 58-60 (1991).
5. Bateman, A. J. *Heredity* **2**, 349-368 (1948).
6. Gwynne, D. T. & Simmons, L. W. *Nature* **346**, 172-174 (1990).
7. Thornhill, R. *Am. Nat.* **110**, 153-163 (1976).
8. Simmons, L. W. & Parker, G. A. *Ethology* **81**, 332-343 (1989).
9. Simmons, L. W. & Bailey, W. J. *Evolution* **44**, 1853-1868 (1990).
10. Bowen, B. J., Codd, C. G. & Gwynne, D. T. *Aust. J. Zool.* **32**, 23-31 (1984).
11. Gwynne, D. T. *Nature* **307**, 361-363 (1984).
12. Gwynne, D. T. *Evolution* **42**, 545-555 (1988).
13. Simmons, L. W. *Behav. Ecol. Sociobiol.* **27**, 43-47 (1990).
14. Gwynne, D. T. *Am. Nat.* **136**, 474-484 (1990).
15. Petrie, M. in *Mate Choice* (ed. Bateson, P.) 167-179 (Cambridge Univ. Press, 1983).
16. Simmons, L. W. & Gwynne, D. T. *Behav. Ecol.* **2**, 276-282 (1991).
17. Wedell, N. & Arak, A. *Behav. Ecol. Sociobiol.* **24**, 117-125 (1989).
18. Clutton-Brock, T. H. *The Evolution of Parental Care* (Princeton Univ. Press, New Jersey, 1991).

|| To whom correspondence should be addressed.

recruitment of facilitation by pre-pulses or by a membrane-permeant cyclic AMP analogue (Fig. 1b). Thus G proteins are not required for voltage-dependent recruitment of facilitation^{7,8}.

The broad spectrum protein kinase inhibitor H-7 (refs 9 and 10) inhibited the recruitment of facilitation by pre-pulses (Fig. 2). In the absence of H-7, a 50-ms pre-pulse to +120 mV recruited ~1,000 pA of facilitation Ca^{2+} current (Fig. 2a, panels 1 and 2). Addition of H-7 did not affect the standard Ca^{2+} current, but abolished the recruitment of facilitation by pre-pulses (Fig.

2a, panels 3 and 4). After the H-7 was washed out, pre-pulses again recruited ~1,000 pA of facilitation Ca^{2+} current (Fig. 2a, panels 5 and 6). In the presence of H-7, BAY K-8644 was still able to activate the facilitation Ca^{2+} current, illustrating that H-7 itself did not directly block the facilitation Ca^{2+} channels¹¹.

Protein kinase inhibitors also inhibited the recruitment of facilitation Ca^{2+} currents by repetitive depolarizations to physiological potentials^{3,4}. Figure 2b shows 15 current traces elicited by depolarizing a chromaffin cell once every 30 seconds (0.033 Hz; left panel). Identical depolarizations delivered once every second (1 Hz) recruited facilitation Ca^{2+} currents resulting in a doubling of the whole-cell current (middle panel). K-252a, another broad spectrum kinase inhibitor^{10,12}, blocked the recruitment of facilitation by 1 Hz depolarization (right panel).

The specific protein kinase A inhibitor peptide PKI 14-24 amide suppressed the response to the D_1 dopamine agonist SKF-38393 (Fig. 2c), but had no effect on the recruitment of facilitation by pre-pulses. Thus recruitment of facilitation by pre-pulses involves phosphorylation by a kinase other than protein kinase A. Furthermore, because the pre-pulse potentials were more positive than the Ca^{2+} equilibrium potential, Ba^{2+} rather than Ca^{2+} was the charge carrier, and the patch pipette contained 10 mM EGTA, it is unlikely that recruitment of facilitation involves a Ca^{2+} -dependent kinase.

Pre-pulse recruitment of facilitation Ca^{2+} current was transient, typically lasting ~60 s (Fig. 3). After the application of okadaic acid, an inhibitor of phosphoprotein phosphatases 1 and 2A (refs 13 and 14), facilitation recruited by a pre-pulse was prolonged indefinitely (Fig. 3a), suggesting that dephosphorylation was involved in the recovery process. Okadaic acid suppresses dephosphorylation, resulting in enhanced phosphorylation. Facilitation current recruited in the presence of okadaic acid was larger than that recruited in its absence.

Replacing ATP in the patch pipette with ATP γS greatly prolonged the facilitation Ca^{2+} currents that were recruited by pre-pulses (Fig. 3b). As dephosphorylation of thio-phosphorylated groups proceeds more slowly than phosphorylated groups^{15,16}, these results also suggest that dephosphorylation is involved in recovery from recruiting facilitation Ca^{2+} currents. Similarly, recruitment of facilitation was slowed in the presence of ATP γS . With ATP in the pipette a single pre-pulse to +120 mV of 100 ms was sufficient to maximally recruit facilitation. But with ATP γS in the pipette, many pre-pulses were required to recruit facilitation maximally (Fig. 3c), suggesting that phosphorylation occurred during the pre-pulse.

Phosphatase 2A is present in large quantities in adrenal chromaffin cells¹⁷ and has been shown to modulate Ca^{2+} channel activity¹⁸. Okadaic acid produced clear effects on facilitation time courses at concentrations as low as 10 nM, indicating that phosphatase 2A may be involved in the dephosphorylation response¹⁴. Figure 4 shows an experiment in which the catalytic subunit of phosphatase 2A (PP-2Ac; ref. 19) was introduced into the cell. Early in the experiment, before a large quantity of PP-2Ac made its way into the cell, pre-pulses recruited large amounts of facilitation Ca^{2+} currents. After more PP-2Ac diffused into the cell, responses to pre-pulses became smaller and shorter (Fig. 4). In experiments in which PP-2Ac was not introduced into the cells, recruitment of facilitation by pre-pulses did not diminish with time. The effects of the PP-2Ac could be reversed by the application of okadaic acid.

Taken together, our results suggest that pre-pulses or repetitive depolarization recruit facilitation Ca^{2+} channels by promoting the phosphorylation of the facilitation Ca^{2+} channel or a closely associated regulatory protein. Recovery from facilitation involves a dephosphorylation step. The rapid and first-order activation of facilitation makes it unlikely that a biochemical cascade is involved in the response. It also suggests that the kinase involved in recruiting facilitation has a high turnover rate, but

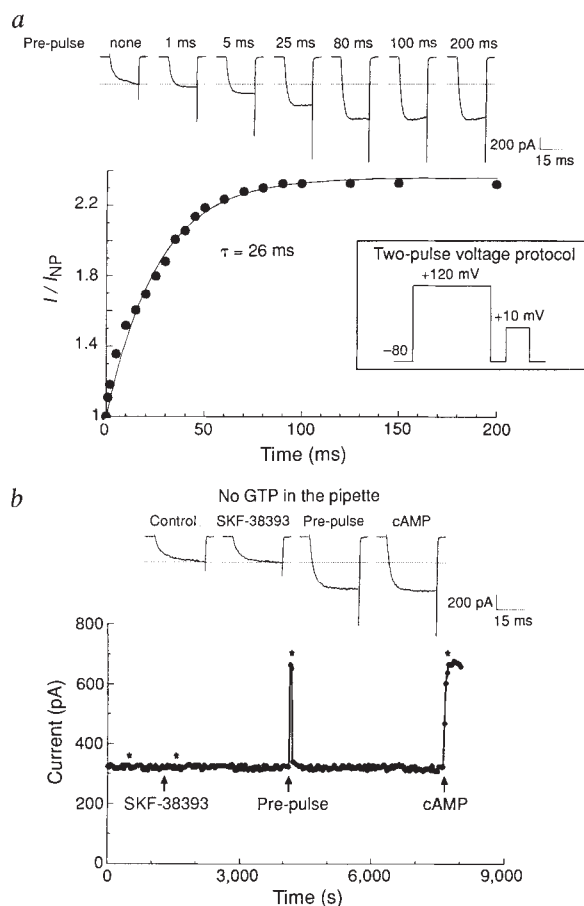
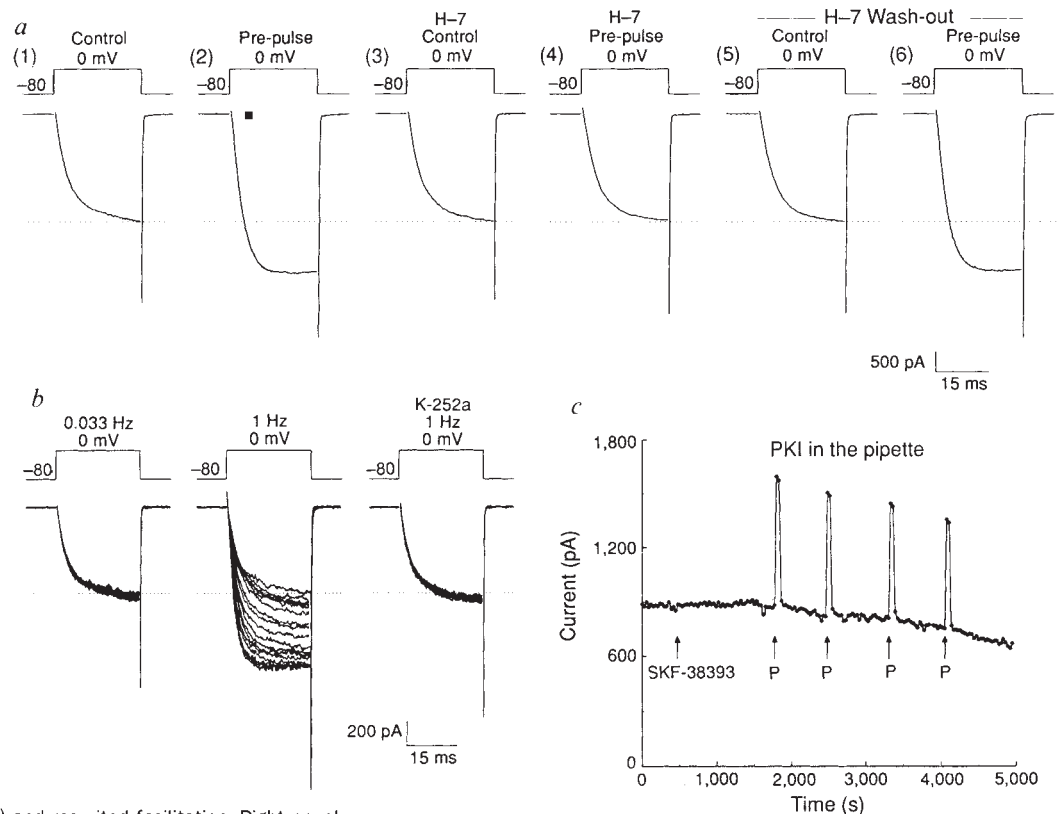


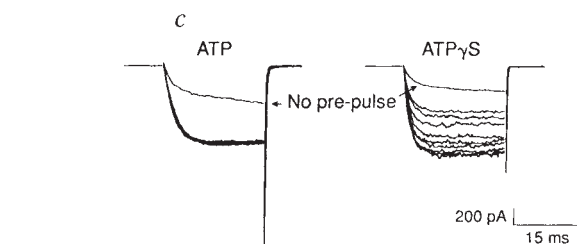
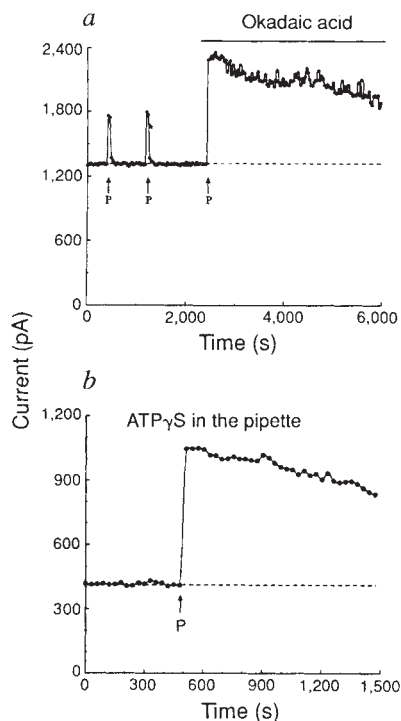
FIG. 1 Recruitment of facilitation Ca^{2+} current in bovine chromaffin cells is rapid at +120 mV. **a**, top, leak-subtracted current traces obtained with the two-pulse voltage protocol shown in the inset. Currents were elicited by a 25-ms test pulse to +10 mV from a holding potential of -80 mV. Pre-pulses to +120 mV of different durations preceded the test pulse. Between the pre- and test pulses, the cell was returned to the holding potential (-80 mV) for 20 ms. Pre-pulse durations were varied from 0–200 ms, as indicated. Only currents from the test pulse are plotted. Bottom, plot of peak current observed after a pre-pulse, normalized by the peak current observed with no pre-pulse, as a function of pre-pulse duration. The data points are fitted by a single exponential with $\tau = 26$ ms. Similar time courses were obtained in 9 cells ($\tau \sim 30$ ms). **b**, GTP was not required for the recruitment of facilitation by pre-pulses. GTP was omitted from the patch pipette. At the point indicated by the first arrow, the D_1 receptor agonist SKF-38393 (10 μM) was added to the bath without effect, indicating that GTP levels were too low to support G protein-mediated responses. At the second arrow a pre-pulse to +120 mV for 100 ms was applied to the cell and facilitation Ca^{2+} current was recruited. At the third arrow the membrane permeant cAMP analogue 8-(4-chlorophenylthio)cAMP (50 μM) was added to the bath. Inset currents were obtained at points labelled with an asterisk. Similar results were obtained in 7 cells.

METHODS. **a**, The pipette solution consisted of (in mM): CsCl (105), Cs-EGTA (10), HEPES (40), MgCl_2 (5), ATP (2), GTP (0.3), adjusted to pH 7.3 with CsOH. In **b**, the internal solution was identical except that GTP was omitted. The external solution contained (in mM): BaCl_2 (10), TEA-Cl (135), HEPES (10) and 1 μM TTX, adjusted to pH 7.3 with TEA-OH. Compounds were applied through a fast-perfusion system.

FIG. 2 Broad range kinase inhibitors such as H-7 or K-252a can inhibit the recruitment of facilitation Ca^{2+} currents by pre-pulses or repetitive depolarizations. *a*, Current traces obtained by depolarizing a chromaffin cell from -80 mV to 0 mV for 25 ms. From left to right: (1) current obtained by the test depolarization; (2) facilitation Ca^{2+} current recruited by the test pulse preceded by a 100 -ms pre-pulse to $+120$ mV (see Fig. 1 legend for voltage protocol); (3) current elicited by a test pulse after H-7 ($200 \mu\text{M}$) was perfused onto the cell; (4) currents recorded after a pre-pulse to $+120$ mV in the presence of H-7; (5 and 6), currents obtained after H-7 was washed out of the bath. Similar results were obtained with either $250 \mu\text{M}$ ($N=8$), $200 \mu\text{M}$ ($N=16$) or $100 \mu\text{M}$ ($N=6$) H-7. *b*, Left panel shows currents elicited by test pulses applied to 0 mV once every 30 s (0.033 Hz). Middle panel shows the currents obtained with the same pulse protocol as panel at left, except that test depolarizations were applied once per second (1 Hz) and recruited facilitation. Right panel shows currents elicited using an identical protocol to that in middle panel (1 Hz), demonstrating that K-252a ($10 \mu\text{M}$) prevented the recruitment of facilitation by voltage. Similar results were obtained with either $10 \mu\text{M}$ ($N=5$) or $20 \mu\text{M}$ ($N=2$) K-252a. *c*, Protein kinase A was not involved in the recruitment of facilitation Ca^{2+} currents by pre-pulses. The specific protein



kinase A-blocker PKI14–24amide ($10 \mu\text{M}$) could suppress the response to D_1 dopamine agonist⁶ (SKF-38393, $10 \mu\text{M}$; first arrow). At locations marked with a 'P', pre-pulses were applied to recruit facilitation. No effect of PKI was observed on pre-pulse recruitment of facilitation. The PKI peptide was in the patch pipette throughout the experiment.



pre-pulse to $+120$ mV for 100 ms was applied to recruit facilitation. Okadaic acid ($1 \mu\text{M}$) was perfused onto the cell as indicated. Similar results were obtained with either $5 \mu\text{M}$ ($N=3$), $2 \mu\text{M}$ ($N=3$), $1 \mu\text{M}$ ($N=18$) or 500 nM ($N=17$) okadaic acid. Some effects were observed at 10 nM okadaic acid ($n=2$). (In about half of the experiments, facilitation Ca^{2+} current disappeared after ~ 30 – 45 min of exposure to okadaic acid. At this time, however, neither pre-pulses nor BAY-K 8644 were able to activate facilitation. As BAY-K 8644 activates facilitation Ca^{2+} channels directly, disappearance of facilitation was probably due to a change in the channels rather than to recovery from okadaic acid treatment). *b*, Replacing the ATP in the patch pipette with $\text{ATP}\gamma\text{S}$ greatly prolonged recruitment of facilitation by pre-pulses. $\text{ATP}\gamma\text{S}$ (2 mM) was present in the patch pipette throughout the experiment. At the point marked with 'P', a pre-pulse was applied to recruit facilitation. Time course of recruiting facilitation can be compared with *a* and Fig. 2*c*. *c*, Recruitment of facilitation was greatly slowed by $\text{ATP}\gamma\text{S}$. Family of curves at left shows test pulse currents elicited every 30 s. ATP was present in the pipette and okadaic acid ($2 \mu\text{M}$) was present in the bath to prevent recovery from facilitation. A pre-pulse to $+120$ mV for 100 ms preceded every test pulse, except for the sweep labelled 'No pre-pulse'. The family of curves at right shows currents obtained with $\text{ATP}\gamma\text{S}$ present in the pipette. (With ATP in the pipette, a 5 -ms pre-pulse to $+120$ mV activated $\sim 18\%$ of maximal facilitation; with $\text{ATP}\gamma\text{S}$ in the pipette, virtually no facilitation was recruited with a 5 -ms pre-pulse (data not shown).)

FIG. 3 Pre-pulses recruited facilitation Ca^{2+} currents reversibly; the phosphoprotein phosphatase inhibitor okadaic acid rendered pre-pulse activation irreversible. Plot of experimental current as a function of time. The chromaffin cell was depolarized to $+10$ mV every 30 s. At locations marked with 'P', a

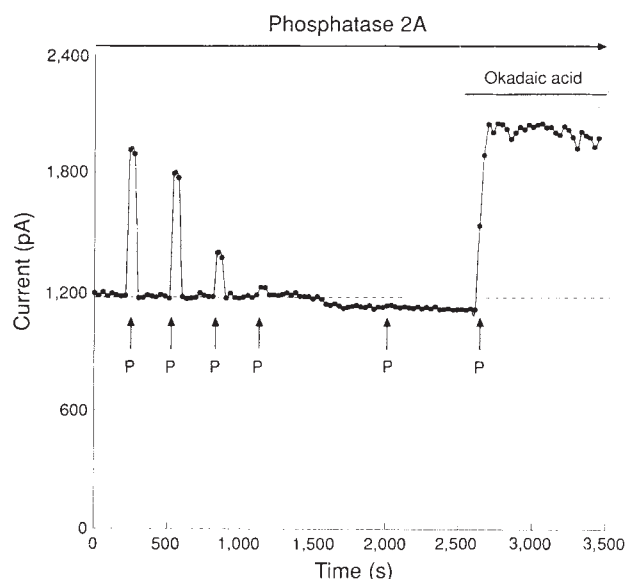


FIG. 4 Introduction of the catalytic subunit of phosphatase 2A (PP-2Ac) into chromaffin cells could prevent pre-pulse recruitment of facilitation Ca^{2+} current. Plot of experimental current as a function of time; the chromaffin cell was depolarized to +10 mV every 30 s. At points marked with 'P' a pre-pulse to +120 mV for 100 ms was applied to recruit facilitation. The pipette internal solution contained ~0.05 unit per ml of PP-2Ac (ref. 20; also 0.5 mM dithiothreitol (DTT), 2 $\mu\text{g ml}^{-1}$ leupeptin and 0.1 mM phenylmethylsulphonylfluoride (PMSF)) which was continuously dialysed into the cell. As the experiment progressed and more PP-2Ac entered the cell, pre-pulses were less effective at recruiting facilitation. Addition of okadaic acid (2 μM) reversed the effects of the PP-2Ac. PP-2Ac was prepared from skeletal muscle according to ref. 20. The enzyme was stored at -20°C in a solution containing 50 mM Tris-HCl, pH 7.0, 0.1 mM EGTA, 0.1% β -mercaptoethanol, 50% glycerol, 1 $\mu\text{g ml}^{-1}$ leupeptin. Before use the enzyme was dialysed for 6 h against a solution containing (in mM): CsCl (105), MgCl_2 (5), HEPES (40), EGTA (10), DTT (0.5), PMSF (0.1), 3 $\mu\text{g ml}^{-1}$ leupeptin, pH 7.3. After dialysis, 2 mM ATP and 0.3 mM GTP were added to the enzyme solution. (In 7 experiments PP-2Ac completely inhibited the recruitment of facilitation by a pre-pulse. In the other 9 experiments, the phosphatase blocked only ~60%–80% of the facilitation Ca^{2+} current. Because we have no independent assay of the amount of PP-2Ac actually introduced into the cells, we assume in these cells dialysis of PP-2Ac was less efficient).

one which is consistent with known kinases^{20,21}. Although protein kinase A can recruit facilitation, it is not involved in recruiting facilitation by pre-pulses. Thus, a second as-yet-unidentified kinase can also recruit facilitation. Because facilitation recruited by pre-pulses is indistinguishable from facilitation recruited by cAMP, the kinase that mediates the recruitment of facilitation by pre-pulses may phosphorylate the same site as protein kinase A. Facilitation is recruited at a lower rate in the presence of ATP γ S than in the presence of ATP, supporting the idea that the Ca^{2+} channel or an associated protein is phosphorylated during the pre-pulse itself. Two mechanisms are plausible. First, depolarization may change the conformation of the facilitation Ca^{2+} channel, making it a better substrate for a kinase. This model is analogous to that developed for the regulation of β -adrenergic receptors. Binding of agonists to the β -adrenergic receptor induces a conformational change in the receptor that makes it a better substrate for phosphorylation by β -adrenergic receptor kinase²². Similarly, large pre-pulses or repetitive depolarizations may produce a conformational change in the facilitation Ca^{2+} channel to make it a better substrate for phosphorylation. Here the voltage-dependence of facilitation and possibly the kinase activity itself would reside in the Ca^{2+} channel^{23,24}. Alternatively, depolarization may directly activate a membrane-associated protein kinase. Distinguishing between these two possibilities will require further studies. □

Received 28 February; accepted 13 May 1992.

1. Fenwick, E. M., Marty, A. & Neher, E. *J. Physiol.* **331**, 599–635 (1982).
2. Hoshi, T., Rothlein, J. & Smith, S. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5871–5875 (1984).
3. Hoshi, T. & Smith, S. J. *J. Neurosci.* **7**, 571–580 (1987).
4. Artalejo, C. R., Dahmer, M. K., Perlman, R. L. & Fox, A. P. *J. Physiol., Lond.* **432**, 681–707 (1990).
5. Artalejo, C. R., Mogul, D. J., Perlman, R. L. & Fox, A. P. *J. Physiol., Lond.* **444**, 213–240 (1991).
6. Artalejo, C. R., Ariano, M. A., Perlman, R. L. & Fox, A. P. *Nature* **348**, 239–242 (1990).
7. Yatani, A. *et al. Science* **238**, 1288–1292 (1987).
8. Yatani, A. & Brown, A. M. *Science* **245**, 71–74 (1989).
9. Hidaka, H., Inagaki, M., Kawamoto, S. & Sasaki, Y. *Biochemistry* **23**, 5036–5041 (1984).
10. Nestler, E. J. & Greengard, G. J. in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects* 4th edn (eds Siegel, G. J., Agranoff, B. W., Albers, R. W. & Molinoff, P. B.) (Raven, New York, 1989).
11. Amador, M. & Dani, J. A. *Neurosci. Lett.* **124**, 251–255 (1991).
12. Kase, H. *et al. Biochem. biophys. Res. Commun.* **142**, 436–440 (1987).
13. Hescheler, J., Mieskes, G., Rüegg, J. C., Takai, A. & Trautwein, W. *Pflügers Arch.* **412**, 248–252 (1988).
14. Cohen, P., Holmes, C. F. B. & Tsukitani, Y. *Trends biochem. Sci.* **15**, 98–102 (1990).
15. Yount, R. G. *Adv. Enzym.* **34**, 1–56 (1975).
16. Gratecos, D. & Fischer, E. H. *Biochem. biophys. Res. Commun.* **58**, 960–967 (1977).
17. Haavik, J. *et al. FEBS. Lett.* **251**, 36–42 (1989).
18. Hescheler, J., Kameyama, M., Trautwein, W., Mieskes, G. & Söling, H.-D. *Eur. J. Biochem.* **165**, 261–266 (1987).
19. Cohen, P. *et al. Meth. Enzym.* **159**, 390–408 (1988).
20. Tessmer, G. W., Skuster, J. R., Tabatabai, L. B. & Graves, D. J. *J. Biol. Chem.* **252**, 5666–5671 (1977).
21. Almers, W. A. *Rev. Physiol.* **52**, 607–624 (1990).
22. Hausdorff, W. P., Caron, M. G. & Lefkowitz, R. L. *FASEB J.* **4**, 2881–2889 (1990).
23. Imagawa, T., Leung, A. T. & Campbell, K. P. *J. Biol. Chem.* **262**, 8333–8339 (1987).
24. Chung, S., Reinhart, P. H., Martin, B. L., Brautigan, D. & Levitan, I. B. *Science* **253**, 56–562 (1991).

ACKNOWLEDGEMENTS. This work was supported by the DGICYT and FIS of Spain (C.R.A.), grants from the Multiple Sclerosis Society (S.R.), the NIH (S.R., A.P.F. and R.L.P.), and the NSF (R.L.P.), and awards from the Klingenstein Fund, McKnight Foundation, Whitaker Foundation and Sprague Foundation (A.P.F.). We thank A. Cahill for preparing cells, R. Miller and H. Fozzard for comments on the manuscript, C. Palfrey for discussion, and A. Sciarina for gifts of nisoldipine and BAY K 8644.

MHC class I alloantigen specificity of Ly-49⁺ IL-2-activated natural killer cells

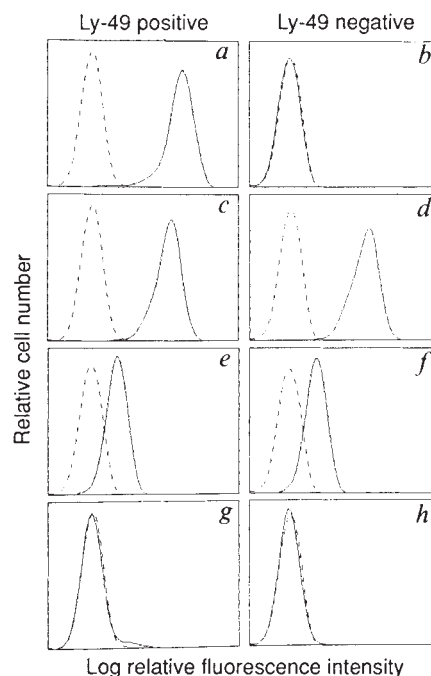
Franz M. Karlhofer, Randall K. Ribaudo* & Wayne M. Yokoyama

The Rosalind Russell Arthritis Research Laboratory, Department of Medicine, University of California, San Francisco, and the Medical Service, San Francisco General Hospital, 1001 Potrero Avenue, San Francisco, California 94110, USA *Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

THE molecular basis of target cell recognition by CD3⁺ natural killer (NK) cells is poorly understood, despite the ability of NK cells to lyse specific tumour cells^{1,2}. In general, target cell major histocompatibility complex (MHC) class I antigen expression correlates with resistance to NK cell-mediated lysis^{3–9}, possibly because NK cell-surface molecules engage MHC class I antigens and consequently deliver inhibitory signals^{3,4}. Natural killer cell allospecificity involves the MHC class I peptide-binding cleft¹⁰, and further understanding of this allospecificity should provide insight into the molecular mechanisms of NK cell recognition. The Ly-49 cell surface molecule is expressed by 20% of CD3⁺ NK cells¹¹ in C57BL/6 mice (H-2^b). Here we show that C57BL/6-derived, interleukin-2-activated NK cells expressing Ly-49 do not lyse target cells displaying H-2^d or H-2^k despite efficient spontaneous lysis by Ly-49⁺ effector cells. This preferential resistance correlates with expression of target cell MHC class I antigens. Transfection and expression of H-2D^d, but not H-2K^d or H-2L^d, renders a susceptible target (H-2^b) resistant to Ly-49⁺ effector cells. The transfected resistance is abrogated by monoclonal antibodies directed against Ly-49 or the $\alpha 1/\alpha 2$ domains of H-2D^d, suggesting that Ly-49 specifically interacts with the peptide-binding domains of the MHC class I alloantigen, H-2D^d. Inasmuch as Ly-49⁺ effector cells cannot be stimulated to lyse H-2D^d targets, our results indicate that NK cells may possess inhibitory receptors that specifically recognize MHC class I antigens.

FIG. 1 Flow cytometric analysis of C57BL/6J IL-2-activated NK cells separated according to surface expression of the Ly-49 antigen. Ly-49⁺ and Ly-49⁻ IL-2-activated NK cells were incubated with monoclonal antibodies (mAbs) specific for Ly-49 (a, b), NK1.1 (c, d), FcγRIII (e, f), or CD3ε (g, h) and then with a fluorescein isothiocyanate (FITC)-conjugated rabbit anti-mouse immunoglobulin antibody (Cappel, Inc.) (a-d, g, h) or FITC-conjugated MAR18.5 (anti-rat κ mAb) (e, f). Specific staining (solid lines) and staining with FITC-conjugated second step reagents alone (dashed lines) are shown. Histograms of unstained cells are identical to those obtained when cells were incubated with second-step reagent alone (not shown). The two effector cell subpopulations (Ly-49⁺ and Ly-49⁻) maintained their phenotypes in culture and were similar with respect to their expression of NK1.1 and murine FcγRIII (CD16). Fewer than 1% of the cells in either population expressed murine CD3 (Fig. 1), CD4, CD8 or surface immunoglobulin (not shown).

METHODS. Ly-49⁺ and Ly-49⁻ IL-2-activated NK cells were prepared. Adherent IL-2-activated, C57BL/6J-derived NK cells were generated by *in vitro* culture in 800 U per ml recombinant human (rh) IL-2 (Hoffman-LaRoche) as described¹⁴, and collected on day 3 with Ca²⁺/Mg²⁺-free PBS containing 0.02% EDTA. IL-2-activated NK cells were depleted of contaminating T cells by incubation with mAbs (0.5% ascites preparations) 53-6.72 (rat anti-CD8; ref. 23), H57-957.1 (hamster anti-TCRαβ; ref. 24), and UC7-13D5 (hamster anti-TCRγδ, provided by J. Bluestone) on ice in HBSS plus 3% FCS for 30 min (2 × 10⁷ cells per ml). Washed cells (10⁷ cells per ml) were incubated with 10 μg ml⁻¹ affinity-purified rabbit anti-mouse immunoglobulin (Cappel), and Low-tox-M rabbit complement (CedarLane) for 45 min at 37 °C. Surviving cells (5 × 10⁵ cells per ml) were washed and expanded with rHIL-2. On day 6, the IL-2-activated NK cells were collected and separated according to surface expression of the Ly-49 antigen by panning with mAb A1 (Ly-49⁺) or depletion with mAb A1 and C' (Ly-49⁻): cells were incubated with mAb A1 (0.5% ascites preparation) on ice for 30 min, washed and panned for 45 min on tissue culture flasks (6 × 10⁶ cells per ml, Ca²⁺/Mg²⁺-free PBS containing 0.02% EDTA, 3% FCS) previously coated with rabbit anti-mouse immunoglobulin (10 μg ml⁻¹ in RPMI). Nonadherent (Ly-49⁻) cells were further depleted of contaminating Ly-49⁺ cells with rabbit C' as described above, then washed and cultured in complete medium containing rHIL-2. Panning flasks containing adherent Ly-49⁺ IL-2-activated NK cells were rinsed three times with HBSS plus 3% FCS to remove all nonadherent cells. Complete medium containing rHIL-2 was then added. After an overnight incubation (day 7), both cell preparations were collected and washed extensively. To avoid potential effects induced by mAb selection, all cells were cultured (10⁶ cells per ml) for an additional 2 days after separation in fresh



complete medium supplemented with rHIL-2. Aliquots of Ly-49⁺ and Ly-49⁻ IL-2-activated NK cells were routinely analysed on day 8 by flow cytometry, with results similar to those shown here. For flow cytometric analysis, cells were incubated with saturating concentrations of mAbs A1 (anti-Ly-49), PK136 (mouse anti-NK1.1; ref. 25), 2.4G2 (rat anti-FcγRIII; ref. 26), or 145-2C11 (hamster anti-CD3ε; ref. 27) as before¹⁴. Flow cytometric analysis was done on a FACScan (Becton Dickinson). Dead cells were excluded by propidium iodide staining and debris was excluded by light scatter parameters; 10,000 events were collected for each histogram. The FITC-anti-mouse immunoglobulin crossreacts with hamster immunoglobulin. There was no evidence of residual mAb A1 by staining with FITC-conjugated anti-immunoglobulin alone.

As Ly-49 seems to be encoded by a multigene family¹² in the mouse NK complex¹³, we have speculated that Ly-49 may be representative of a family of NK cell surface molecules that recognize polymorphic ligands¹⁴. To address this possibility, we prepared Ly-49⁺ and Ly-49⁻ effector cells and determined their target cell specificity. Except for Ly-49 expression, the CD3⁻ effector cell subsets were equivalent with respect to expression of NK cell surface molecules (Fig. 1) and ability to lyse certain targets (Table 1).

The Ly-49⁺ effector cells, however, could not lyse spontaneously a panel of other tumour cells that were lysed by Ly-49⁻ effector cells (Table 1). The preferentially resistant targets could not be lysed by Ly-49⁺ effector cells even by other mechanisms (such as antibody-dependent cellular cytotoxicity, anti-NK1.1 monoclonal antibody-induced redirected lysis¹⁴ or lectin-induced lysis) (data not shown). Preferential target cell resistance correlated with target cell H-2 haplotype (Table 1). Target cells homozygous or heterozygous (with H-2^b) for the H-2^d or H-2^k haplotypes were resistant to lysis by Ly-49⁺ effector cells. The only exceptions to this correlation were YAC-1 cells (H-2^a, including H-2K^k, H-2D^d), which were very sensitive to lysis by effector cells, and WEHI-3 and BB88 (H-2^d) cells, which were resistant to both cell subsets. Other factors may influence the susceptibility of these exceptional targets to both Ly-49⁺ and Ly-49⁻ effector cells. These results suggested that an MHC antigen of the H-2^k or H-2^d haplotype is a target cell ligand for Ly-49. Furthermore, R1.1 cells (H-2^k) were preferentially resistant to lysis by Ly-49⁺ effector cells but derivative R1E/TL8x.1 cells, which do not express MHC class I antigens (β₂-microglobulin gene mutation¹⁵), were susceptible to lysis

by both effector cell subsets (Table 1). The H-2^k-derived cell line LBRM-33-1A5 was susceptible to lysis by both effector cell subsets (Table 1) but it did not express MHC class I antigens. Thus, target cell surface expression of MHC class I antigens of certain haplotypes correlated with resistance to lysis by Ly-49⁺ effector cells.

To explore further the relation between specific MHC class I antigens and target cell resistance to Ly-49⁺ effector cells, we transfected the C1498 cell line (H-2^b), which was susceptible to lysis by Ly-49⁺ and Ly-49⁻ effector cells (Table 1), with complementary DNAs encoding H-2D^d, H-2K^d, H-2L^d, or with empty vector DNA (Fig. 2A). Only transfected C1498 cells expressing H-2D^d became resistant to lysis by Ly-49⁺ interleukin-2 (IL-2)-activated NK cells (Fig. 2B). These transfectants remained susceptible to spontaneous lysis by Ly-49⁻ effector cells, indicating that the gene-transferred resistant phenotype was specific for Ly-49⁺ effector cells.

Transfected resistance to Ly-49⁺ IL-2-activated NK cell-mediated lysis was reversed by anti-Ly-49 monoclonal antibody, supporting a specific role for Ly-49 in recognizing H-2D^d on target cells (Fig. 3A). As we used the F(ab')₂ anti-Ly-49 antibody and as C1498 does not express Fc receptors, it is unlikely that anti-Ly-49 induced redirected lysis¹⁴. Moreover, anti-Ly-49 does not induce lysis of Daudi cells which are susceptible to anti-NK1.1 antibody-induced lysis by these effector cells (data not shown; ref. 14). Blocking studies with anti-H-2D^d-specific antibodies (Fig. 3B) demonstrated that an anti-H-2D^d-α1/α2 domain-specific antibody allowed Ly-49⁺ NK cells to lyse the H-2D^d-transfected target. In contrast, an anti-H-2D^d-α3 domain-specific antibody did not affect lysis. Similar results

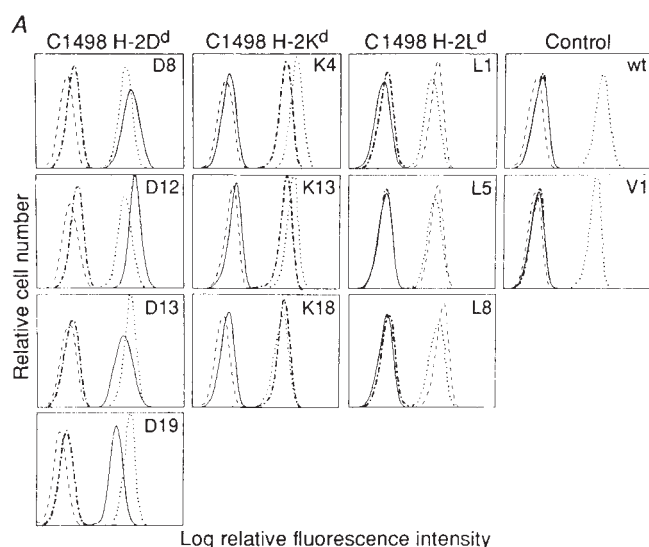


FIG. 2 A, Flow cytometric analysis of C1498 cells (H-2^b) and C1498 cells transfected with H-2D^d, H-2K^d, H-2L^d or vector DNA alone. Cells were incubated with mAb 34-2-12S (anti-H-2D^d, solid lines; ref. 28), mAb SF1-1.1.1 (anti-H-2K^d, heavy dotted line; ref. 29), mAb 30-5-7S (anti-H-2L^d, dashed lines; ref. 28), or mAb AF6-88.5.3 (anti-H-2K^b, fine dotted line; ref. 30), followed by FITC-conjugated goat anti-mouse immunoglobulin antibody. For clarity, background staining with FITC-conjugated second step reagent alone is not shown but was identical to the lowest level staining. Before any testing in cytotoxicity experiments, these cell lines were chosen from a larger panel of transfectants because of high levels of comparable expression of the transfected gene product. B, C1498 cells (H-2^b) transfected with H-2D^d were preferentially resistant to lysis by Ly-49 positive effector cells whereas transfection of the H-2K^d or H-2L^d cDNA did not influence spontaneous killing by Ly-49⁺ effector cells. For each cDNA transfection, 3 independent cell lines were analysed: C1498-H-2D^d: D8 (a), D12 (b), D13 (c); C1498-H-2K^d: K4 (d), K13 (e), K18 (f); C1498-H-2L^d: L1 (g), L5 (h), L8 (i); as well as the following controls: C1498 V.1 (transfected with vector alone) (j), wild-type C1498 (k), and YAC-1 target cells (l). The cell lines assayed for spontaneous lysis by Ly-49⁺ (filled circles) or Ly-49⁻ (circles) effector cells.

METHODS. The H-2D^d cDNA was obtained by amplification by polymerase chain reaction (PCR) of cDNA derived from DBA/2J mouse liver with primers specific for the published H-2D^d gene product. Sequence analysis of the H-2D^d cDNA showed identity with published sequences except for a single codon (Arg 218 for Gln 218) alteration in the α_3 domain, and a deletion (amino acids 316–340) in the cytoplasmic tail (data not shown). The cDNA

were obtained with the preferentially resistant H-2D^d target cells (Table 1; and other data not shown). These findings suggest that Ly-49 on activated NK cells interacts with the $\alpha 1/\alpha 2$ domains of the H-2D^d molecule.

Our studies provide a basis for further understanding MHC class I-associated target cell resistance to NK cells and NK cell allospecificity. As Ly-49⁺ IL-2-activated NK cells were unable to lyse H-2D^d-expressing targets by several distinct mechanisms, interaction of Ly-49 with an allo-MHC class I antigen (H-2D^d) globally and negatively affected NK cell stimulation. The anti-Ly-49 and anti-H-2D^d antibodies appeared to block this interaction, allowing lysis to occur. Taken together, these data suggest that engagement of Ly-49 by H-2D^d delivers an inhibitory signal to activated NK cells. As H-2D^d expression is correlated with target cell resistance to NK cell-mediated lysis by multiple stimuli (spontaneous killing, antibody-dependent cellular cytotoxicity, monoclonal antibody-induced redirected lysis, and lectin-induced lysis), it is less likely that MHC class I sterically hinders or masks all target cell molecules responsible for triggering lysis by these distinct stimuli^{3,4,10}. Our studies do not, however, exclude this masking hypothesis.

NK cell allospecificity is presumably also regulated by other allotypic receptors that initiate rather than inhibit cytotoxicity.

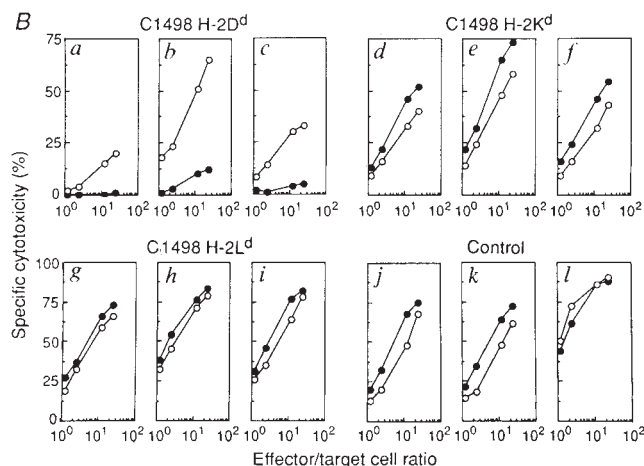
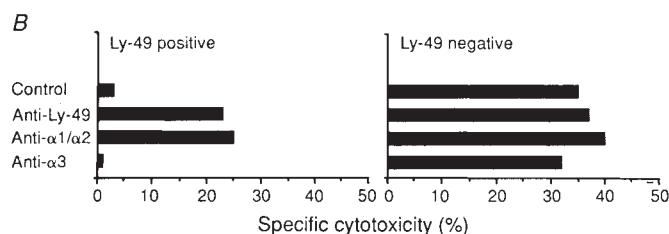
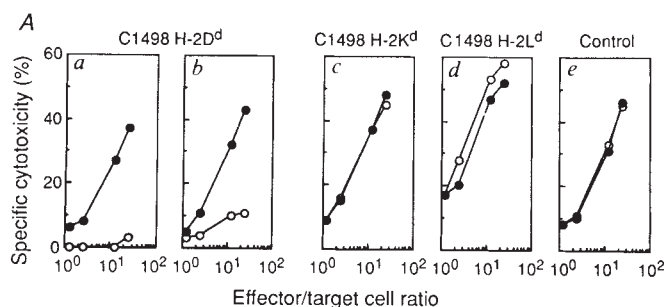


FIG. 3 A, Resistance to Ly-49⁺ IL-2-activated NK cells transfected with the H-2D^d gene is specific for Ly-49. Cytotoxicity by Ly-49⁺ effector cells against C1498H-2D^d: D12 (a), D19 (b), C1498-H-2K^d: K13 (c), C1498H-2L^d: L5 (d) or C1498 V.1 (e) in the presence (filled circles) or absence (circles) of F(ab')₂ anti-Ly-49 mAb (A1) is shown. Incubation with the anti-Ly-49 mAb enabled Ly-49⁺ effector cells to lyse transfected resistant tumour cells to a level that was comparable to spontaneous lysis of the transfected susceptible targets. This was also comparable to lysis of the same targets by Ly-49⁻ effector cells (not shown). The effect of mAb A1 was specific for Ly-49⁺ effector cells. An isotype-matched control mAb did not affect lysis of these targets by Ly-49⁺ effector cells and mAb A1 did not affect lysis of any targets by Ly-49⁻ effector cells (not shown). Similar results with mAb A1 were obtained for all targets that were preferentially resistant to Ly-49⁺ effector cells (not shown). The functional effects of mAb A1 were not due to redirected lysis, a mechanism requiring that an Fc receptor on the target cell bind to the Fc portion of the stimulating antibody^{14,34}. First, mAb A1 allowed Ly-49⁺ effector cells to lyse resistant target cells that do not express Fc receptors (C1498-H-2D^d cell lines, and L5178Y-R, R1.1, BW5147.G1.4; not shown). Second, F(ab')₂ fragments of mAb A1 were as effective as the intact antibody in allowing lysis. Third, Daudi cells, which were susceptible to mAb-induced redirected lysis by Ly-49⁺ NK cells, using mAbs directed against the NK1.1, VEA, and Ly-6 antigens (see above), were not lysed by Ly-49⁺ NK cells incubated with mAb A1 (not shown). B, The

TABLE 1 Lysis of murine tumour cells by Ly-49⁺ and Ly-49⁻ IL-2-activated NK cells

Cell line	Type	H-2	(n)	% Specific cytotoxicity by					
				Ly-49 ⁺ NK cells			Ly-49 ⁻ NK cells		
				25	12.5	2.5	25	12.5	2.5
YAC-1	T cell	a	(17)	76	61	32	81	69	36
1C-21	Macrophage	b	(1)	30	22	8	39	34	11
B16S	Melanoma	b	(1)	81	70	34	73	62	32
C1498	Lymphoma	b	(5)	53	41	15	40	34	14
WEHI-3	Macrophage	d	(1)	2	1	2	2	2	2
BB88	Leukaemia	d	(1)	3	1	0	4	5	3
J774	Macrophage	d	(2)	0	0	0	76	61	23
P388D1	Macrophage	d	(7)	3	2	0	62	44	25
RAW264.7	Macrophage	d	(2)	7	6	5	75	62	29
PU5-1R	Macrophage	d	(1)	0	0	0	53	36	13
WEHI 265.1	Macrophage	d	(4)	2	0	0	55	38	20
WR19M.1	Macrophage	d	(2)	14	9	1	33	30	14
RAW309Cr.1	Macrophage	d/b	(1)	2	0	0	15	11	5
SP2/O	B cell	d	(1)	8	0	0	51	33	10
LB27.4	B cell	d/b	(1)	0	0	0	20	12	7
L5178Y-R	Leukaemia	d	(4)	8	2	2	65	44	19
T27A	Leukaemia	d	(1)	0	0	0	18	17	8
D2N	Leukaemia	d	(1)	8	5	2	47	31	13
DAP3	Fibroblast	k	(1)	3	3	2	34	27	12
BW5147.G.1.4	T cell	k	(2)	7	6	5	30	25	12
R1.1	T cell	k	(2)	3	2	2	24	18	10
R1E/TL8x.1	T cell	k ⁻	(1)	30	26	15	27	22	12
LBRM-33-1A5	Lymphoma	k ⁻	(1)	28	26	12	31	28	11

Lysis of target cells was determined in triplicate at the indicated effector:target (E:T) cell ratios (25:1, 12.5:1, and 2.5:1) in standard ⁵¹Cr-release cytotoxicity assays¹⁴ at 37 °C for 4 h with day-9 IL-2-activated NK cells against the indicated tumour cells. Results are expressed as per cent specific cytotoxicity as described¹⁴. To obtain enough effector cells and to avoid any functional effect produced by anti-Ly-49 antibody based separation, we prepared Ly-49⁺ and Ly-49⁻ effector cells from IL-2-activated NK cells as detailed in Fig. 1 legend. (Inasmuch as monoclonal antibody A1 (anti-Ly-49; ref. 21) was required to prepare the NK cell subsets, and as this antibody had functional effects on Ly-49⁺ effector cells (see later), results were uninterpretable if these effector cell subsets were prepared from freshly isolated NK cells.) The NK cell subsets were prepared in 17 independent experiments. In every experiment, lysis of the prototypical murine NK cell-sensitive target, YAC-1, was assayed. Results were comparable in each experiment, showing that both effector cell subsets lyse this target equivalently. Other targets were tested for lysis by the effector cell subsets in *n* experiments; results were comparable but are shown for only one representative experiment so that the data could be organized with respect to target-cell resistance and H-2 haplotypes. Targets were obtained from the American Type Culture Collection; origins of the cell lines are referenced in their catalogue. The cell type and H-2 haplotype were verified by examination of primary references and by flow cytometry using monoclonal antibodies specific for private epitopes on H-2 molecules. R1E/TL8x.1 (ref. 15) and LBRM-33-1A5 lack MHC class I expression by flow cytometric analysis with the anti-H-2 class antibody M1/42.3.9.8 (ref. 22) (data not shown). The absence of class I MHC expression on LBRM-33-1A5 cells has not been reported previously. Some targets (for example, YAC-1, 1C-21, B16S and C1498) were killed equally well by both effector cell subsets and others (such as BB88 and WEHI-3) were not killed by either subset. Interestingly, several targets were resistant to Ly-49⁺ effector cells. Targets preferentially resistant to Ly-49⁺ effector cells could not be lysed by these effector cells even by the mechanism of antibody-dependent cellular cytotoxicity (ADCC; not shown). Moreover, resistant targets that expressed Fc receptors could not be killed by redirected lysis with antibodies directed against the NK1.1, VEA, or Ly-6 antigens¹⁴, or by lectin (concanavalin A)-induced lysis (not shown), even though Ly-49⁺ and Ly-49⁻ effector cells were equivalent with respect to their ability to mediate ADCC, monoclonal antibody-induced redirected lysis, or lectin-induced lysis of the human target, Daudi (data not shown).



$\alpha 1/\alpha 2$ -domains of H-2D^d are involved in the H-2D^d gene-transferred resistance to Ly-49⁺ effector cells. F(ab')₂ fragments of the indicated mAbs were incubated with the Ly-49⁺ or Ly-49⁻ effector cells and the C1498.D12 target cell.

METHODS. Ly-49⁺ IL-2-activated NK cells were prepared and used in cytotoxicity experiments at varying E:T ratios as described in Table 1. The mAbs were affinity purified by protein A-Sepharose chromatography according to standard procedures. F(ab')₂ fragments were generated by pepsin

digestion and then depleted of undigested mAb and Fc fragments by passage over a protein A-Sepharose column. Purity was determined by 7% SDS-PAGE and biological activity was confirmed by flow cytometry. The purified mAbs and their F(ab')₂ fragments bound to the effector cells equivalently. F(ab')₂ fragments of mAbs A1, 34-5-8S (anti-H-2D^d- $\alpha 1/\alpha 2$; ref. 28), and 34-2-12S (anti- $\alpha 3$; ref. 28) were added to the cytotoxicity assay at 25 μ g ml⁻¹, 25:1 E:T ratio, as indicated. F(ab')₂ fragments of the mAbs were used to avoid the possibility of redirected lysis (anti-Ly-49) or ADCC (anti-H-2D^d).

Ly-49 is genetically linked to *NK1.1* in the mouse NK complex^{11,13}. *NK1.1* encodes an allotypic molecule expressed by all murine NK cells and that activates, rather than inhibits, cytotoxicity^{11,14}. But the engagement of Ly-49 has a dominant negative effect over stimulation through *NK1.1*. Such negative effects are seen in B-cell signalling, where cross linking of FcγRII-B1 inhibits anti-immunoglobulin-induced stimulation¹⁶. Thus, allospecificity of human CD3⁺ NK cell clones^{17,18} may be regulated largely by inhibitory receptors, similar to Ly-49 (ref. 19), and their putative MHC ligands²⁰.

Our blocking studies with monoclonal antibodies implicate the peptide binding domain of H-2D^d in MHC class I-associated target cell resistance to activated NK cells. This is consistent with transfection studies demonstrating that a single amino-acid mutation (residue 74) in the α1 helix of HLA-A2 converted target cells to an NK cell-resistant phenotype¹⁰. Inasmuch as this alteration provides access to a side pocket in the peptide-binding cleft, MHC class I molecules may present peptides to NK cells as well as to T cells.

The physiological relevance of NK cell allospecificity remains to be determined. Molecular dissection of other allo-recognition phenomena, namely tissue rejection, and T-cell allospecificity, have provided insight into the normal physiological operations of these complex biological systems. Likewise, further investigation of NK cell allospecificity should help us understand NK cell recognition. This could explain an apparent paradox in our study; Ly-49⁺ NK cells lyse syngeneic tumour cells but not normal syngeneic cells. Perhaps NK cell (inhibitory) allospecificity mimics NK cell recognition of self-MHC class I molecules bearing specific peptides that may be present in normal cells but absent in syngeneic tumour cells. Whatever the explanation, our results have implications for the regulation of T-cell and NK-cell activity *in vivo* because target cell MHC class I antigens seem to influence both types of cytotoxic lymphocytes but with opposite functional consequences. Finally, this raises the possibility that certain associations of MHC class I antigens with human diseases may be related to the effect of MHC class I molecules on NK cells. □

Received 2 March; accepted 11 May 1992.

- Herberman, R. B. *NK Cells and Other Natural Effector Cells* (Academic, New York, 1982).
- Trinchieri, G. *Adv. Immun.* **47**, 187–376 (1989).
- Ljunggren, H.-G. & Kärre, K. *Immun. Today* **11**, 237–244 (1990).
- Kärre, K. in *Mechanisms of Cytotoxicity by NK Cells* (eds Herberman, R. B. & Callewaert, D. M.) 81–92 (Academic, New York, 1985).
- Quillet, A. *et al. J. Immun.* **141**, 17–20 (1988).
- Sturmhofel, K. & Hammerling, G. J. *Eur. J. Immun.* **20**, 171–177 (1990).
- Liao, N.-S., Bix, M., Zijlstra, M., Jaenish, R. & Raulet, D. *Science* **253**, 199–202 (1991).
- Hoglund, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 10332–10336 (1991).
- Bix, M. *et al. Nature* **349**, 329–331 (1991).
- Storkus, W. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 5989–5992 (1991).
- Yokoyama, W. M., Kehn, P. J., Cohen, D. I. & Shevach, E. M. *J. Immun.* **145**, 2353–2358 (1990).
- Yokoyama, W. M., Jacobs, L. B., Kanagawa, O., Shevach, E. M. & Cohen, D. I. *J. Immun.* **143**, 1379–1386 (1989).
- Yokoyama, W. M. *et al. J. Immun.* **147**, 3229–3236 (1991).
- Karlhofer, F. M. & Yokoyama, W. M. *J. Immun.* **146**, 3662–3673 (1991).
- Hyman, R. & Stallings, V. *Immunogenetics* **3**, 75–84 (1976).
- Hunziker, W., Koch, T., Whitney, J. A. & Mellman, I. *Nature* **345**, 628–632 (1990).
- Moretta, A. *et al. J. exp. med.* **172**, 1589–1598 (1990).
- Suzuki, N. *et al. J. exp. med.* **172**, 457–462 (1990).
- Houchins, J. P., Yabe, T., McSherry, C. & Bach, F. H. *J. exp. med.* **173**, 1017–1020 (1991).

- Ciccone, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 9794–9797 (1990).
- Nagasawa, R. *et al. J. Immun.* **138**, 815–824 (1987).
- Springer, T. A. in *Monoclonal Antibodies* (eds Kennett, R. *et al.*) 185–217 (Plenum, New York, 1980).
- Ledbetter, J. A. & Herzenberg, L. A. *Immunol. Rev.* **47**, 63–90 (1979).
- Kubo, R. T., Born, W., Kappler, J. W., Marrack, P. & Pigeon, M. *J. Immun.* **142**, 2736–2742 (1989).
- Koo, G. C. & Peppard, J. R. *Hybridoma* **3**, 301–303 (1984).
- Perussia, B. *et al. J. exp. med.* **170**, 73–86 (1989).
- Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374–1378 (1987).
- Evans, G. A., Margulies, D. H., Shykind, B., Seidman, J. G. & Ozato, K. *Nature* **300**, 755–757 (1982).
- Rubocki, R. J. *et al. J. Immun.* **146**, 2352–2357 (1991).
- Loken, M. R. & Stall, A. M. *J. Immunol. Meth.* **50**, R85–R112 (1982).
- Miller, J. & Germain, R. G. *J. exp. med.* **164**, 1478–1489 (1986).
- Gunning, P., Leavitt, J., Muscat, G., Ng, S.-Y. & Kedes, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4831–4835 (1987).
- Cochet, M. *et al. Immunogenetics* **24**, 267–274 (1986).
- Leo, O. *et al. J. Immun.* **137**, 3874–3880 (1986).

ACKNOWLEDGEMENTS. We thank D. Margulies for discussion, and W. E. Seaman and I. M. Goldstein for comments on the manuscript. F.M.K. was supported by the Max Kade Foundation. R.K.R. was supported by the Cancer Research Institute and the National Multiple Sclerosis Society. W.M.Y. is a Scholar of the Rosalind Russell Arthritis Research Center. This work was supported by grants from the NIH to W.M.Y., and by the USC Multipurpose Arthritis Center grant.

Rapamycin selectively inhibits interleukin-2 activation of p70 S6 kinase

Calvin J. Kuo, Jongkyeong Chung*, David F. Fiorentino, W. Michael Flanagan, John Blenis* & Gerald R. Crabtree

Howard Hughes Medical Institute, Unit in Molecular and Genetic Medicine, Beckman Center, Stanford University School of Medicine, Stanford, California 94305-5425, USA

* Department of Cellular and Molecular Physiology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

THE macrolide rapamycin induces cell cycle G1 arrest in yeast and in mammalian cells^{1–3}, which suggests that an evolutionarily conserved, rapamycin-sensitive pathway may regulate entry into S phase. In mammals, rapamycin inhibits interleukin-2 receptor-induced S phase entry and subsequent T-cell proliferation^{4–6}, resulting in immunosuppression. Here we show that interleukin-2 selectively stimulates the phosphorylation and activation of p70 S6 kinase but not the *erk*-encoded MAP kinases and *rsk*-encoded S6 kinases^{7,8}. Rapamycin completely and rapidly inhibits interleukin-2-induced phosphorylation and activation of p70 S6 kinase at concentrations comparable to those blocking S phase entry of T cells (0.05–0.2 nM). The structurally related macrolide FK506 competitively antagonizes the actions of rapamycin, indicating that these effects are mediated by FKBP, which binds the transition-state mimic structure common to both rapamycin and FK506 (refs

4, 6, 9–11). The selective blockade of the p70 S6 kinase activation cascade by the rapamycin-FKBP complex implicates this signalling pathway in the regulation of T cell entry into S phase.

We investigated the activation of three serine/threonine protein kinases, *erk*-encoded MAP kinase, *rsk*-encoded S6 kinase (RSK) and p70 S6 kinase, by interleukin-2 (IL-2) in the IL-2-dependent T and B lymphoid cell lines D10 (ref. 5) and BA/F3 (ref. 12). By immune complex kinase assay, IL-2 treatment (30 min) increased p70 S6 kinase activity 8–10-fold in quiescent D10 and BA/F3 cells (Fig. 1a). Stimulation of p70 S6 kinase was detectable within 10 min of IL-2 treatment (Fig. 1b), consistent with the kinetics of p70 S6 kinase activation by other mitogens in non-lymphoid cells^{13,14}, and occurred in serum-free medium (data not shown), in agreement with the serum-independence of IL-2 action¹⁵. Although all three kinases are stimulated by epidermal, platelet-derived and fibroblast growth factors in 3T3 cells^{13,14,16}, IL-2 did not significantly alter the activity of RSK and MAP kinases over the 1-hour time course (Fig. 1a and b, and data not shown), suggesting that the IL-2 receptor may be selectively coupled to p70 S6 kinase.

As rapamycin antagonizes IL-2-stimulated T cell S phase entry^{4,5}, we evaluated the rapamycin sensitivity of p70 S6 kinase. Rapamycin (at 2 ng ml⁻¹) completely inhibited IL-2-stimulated p70 S6 kinase activity in both D10 and BA/F3 cells, whereas the structurally related macrolide, FK506, which does not interfere with IL-2 receptor signalling^{4,5}, was ineffective (Fig. 2a). Rapamycin and FK506 bind to FKBP proteins through a common transition-state mimic structure, an observation that is thought to underlie their mutual antagonism of different phases of progression through the cell cycle in T cells⁴. This mutual

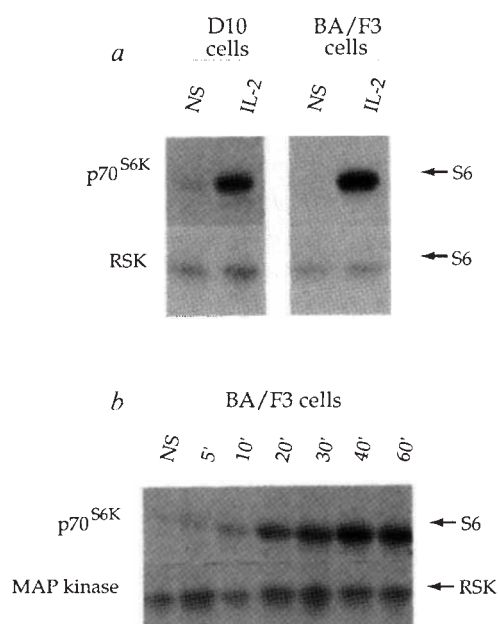


FIG. 1 Interleukin-2 receptor (IL-2R) selectively couples to p70 S6 kinase (p70^{S6K}). *a*, IL-2 activates p70 S6 kinase but not RSK in D10 and BA/F3 cells. NS, non-stimulated. *b*, Time course of IL-2-stimulated p70 S6 kinase activity in BA/F3 cells.

METHODS. D10 and BA/F3 cells were cultured at $<1 \times 10^6$ cells per ml, at 37 °C/7% CO₂ in RPMI with 10% FCS, 100 U per ml penicillin, 100 $\mu\text{g ml}^{-1}$ streptomycin, 50 μM β -mercaptoethanol, supplemented with IL-2 (DNAX), and for BA/F3 cells, 400 $\mu\text{g ml}^{-1}$ G418 (GIBCO) (active drug). Cells were washed with PBS and placed into medium without IL-2 for 15 h (BA/F3) or 24 h (D10) before stimulation with IL-2 for 30 min (*a*) or the indicated times (*b*). Cells were subsequently washed with cold PBS and lysed in (10 mM HEPES, pH 7.8, 15 mM KCl, 1 mM EDTA, 2 mM EGTA, 10% glycerol, 1 mM PMSF, 1 mM DTT, 10 $\mu\text{g ml}^{-1}$ leupeptin, 10 $\mu\text{g ml}^{-1}$ antipain, 10 $\mu\text{g ml}^{-1}$ pepstatin A, 0.1 mM Na₃VO₄, 20 mM β -glycerol phosphate, 0.2% NP-40); all manipulations were carried out at 4 °C. Nuclei and membranes were pelleted by microfuge (15 min) and the supernatants (crude cytoplasmic extracts) stored at -80 °C. Equal amounts of protein (100–400 μg) were assayed for p70 S6 kinase and RSK by immune complex assay using 40S ribosomal subunits (containing ribosomal protein S6) as substrate and [γ -³²P]ATP, as described elsewhere¹⁴. We have occasionally observed slight (two-fold) elevation of RSK activity in response to IL-2 (data not shown). MAP kinase activity was assayed by immune complex assay using recombinant RSK as substrate and [γ -³²P]ATP, as previously described^{13,14}. Phosphorylated substrates were resolved by 12% SDS-PAGE for S6 kinase assay and 7.5% SDS-PAGE for MAP kinase assay.

antagonism presumably reflects the competitive formation of FKBP-rapamycin and FKBP-FK506 complexes with distinct ternary effector proteins, depending on the immunosuppressant in excess, and redirection of the complex to interfere with either the G0→G1 (FK506) or G1→S transitions (rapamycin). Accordingly, a 2,000-fold molar excess of FK506 (4 $\mu\text{g ml}^{-1}$) reversed rapamycin inhibition of p70 S6 kinase activity in D10 and BA/F3 cells, indicating that an FKBP that binds the structural element shared by rapamycin and FK506 is the receptor for these effects (Fig. 2a). Rapamycin specifically inhibited Jurkat cell basal and phorbol myristate acetate (PMA)/ionomycin-stimulated p70 S6 kinase activity but did not alter PMA/ionomycin-stimulated RSK activity (Fig. 2b), further arguing against a non-selective signal transduction block. Additionally, rapamycin produced only minimal changes in the phosphoprotein profile or the phosphotyrosine profile of IL-2 treated BA/F3 cells or serum-treated fibroblasts (data not shown).

The phosphorylation state of p70 S6 kinase is controlled by an interplay between as-yet uncharacterized upstream kinases and phosphatases, with serine phosphorylation of p70 S6 kinase being accompanied by decreased electrophoretic mobility and increased kinase activity^{17,23}. IL-2 treatment of D10 and BA/F3

cells converted unphosphorylated p70 S6 kinase into characteristically phosphorylated, more slowly migrating species on western blotting (Fig. 3a), whereas potato acid phosphatase treatment of phosphorylated p70 S6 kinase produced only the fastest-migrating, dephosphorylated form (data not shown). Moreover, IL-2-induced phosphorylation of p70 S6 kinase was entirely inhibited by rapamycin but not FK506, with reversal of the rapamycin effect observed with 2,000-fold excess of FK506 in both cell types (Fig. 3a), mirroring p70 S6 kinase activity (Fig. 2a). Treatment of D10 cells with increasing doses of rapamycin revealed that p70 S6 kinase enzymatic activity and p70 S6 kinase phosphorylation were blocked concomitantly (Fig. 3b), with the appearance of phosphorylated p70 S6 kinase species correlating strongly with enzymatic activity. The activity of p70 S6 kinase was more sensitive to rapamycin (concentration giving 50% inhibition, $\text{IC}_{50} = 0.05 \text{ ng ml}^{-1}$) than IL-2-stimulated [³H]thymidine (TdR) incorporation determined in parallel ($\text{IC}_{50} = 0.2 \text{ ng ml}^{-1}$) (Fig. 3b, graph). Rapamycin interacts with FKBP₁₂ with $K_d \approx 0.2 \text{ ng ml}^{-1}$ (ref. 4), a concentration at which p70 S6 kinase activity is completely inhibited (Fig. 3b). These observations are consistent with a gain-of-function model in which rapamycin exerts its biological effects despite association

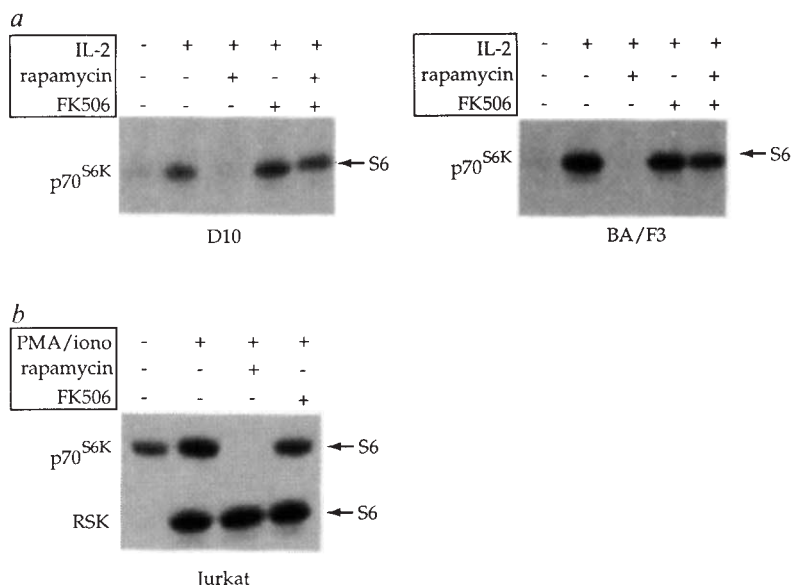


FIG. 2 Rapamycin selectively blocks induction of IL-2-stimulated p70 S6 kinase activity by interaction with an FKBP. *a*, Rapamycin inhibits IL-2 activation of p70 S6 kinase, with reversal of rapamycin inhibition by FK506. *b*, Rapamycin does not antagonize the RSK signalling pathway.

METHODS. D10 and BA/F3 cells were cultured and stimulated with IL-2 for 30 min as for Fig. 1. Jurkat T lymphocytes were grown in RPMI with 10% FCS, 100 U per ml penicillin, 100 $\mu\text{g ml}^{-1}$ streptomycin, and stimulated for 30 min with 2 μM ionomycin and 20 ng ml⁻¹ PMA. Where appropriate, cells were pre-incubated with rapamycin (2 ng ml⁻¹), or FK506 (4 $\mu\text{g ml}^{-1}$) for 5 min preceding stimulation. In *b*, FK506 was used at 2 ng ml⁻¹, as appropriate. Immune complex kinase assays for p70 S6 kinase and RSK are described in Fig. 1 legend and used 40S ribosomal subunits (containing S6) as substrate.

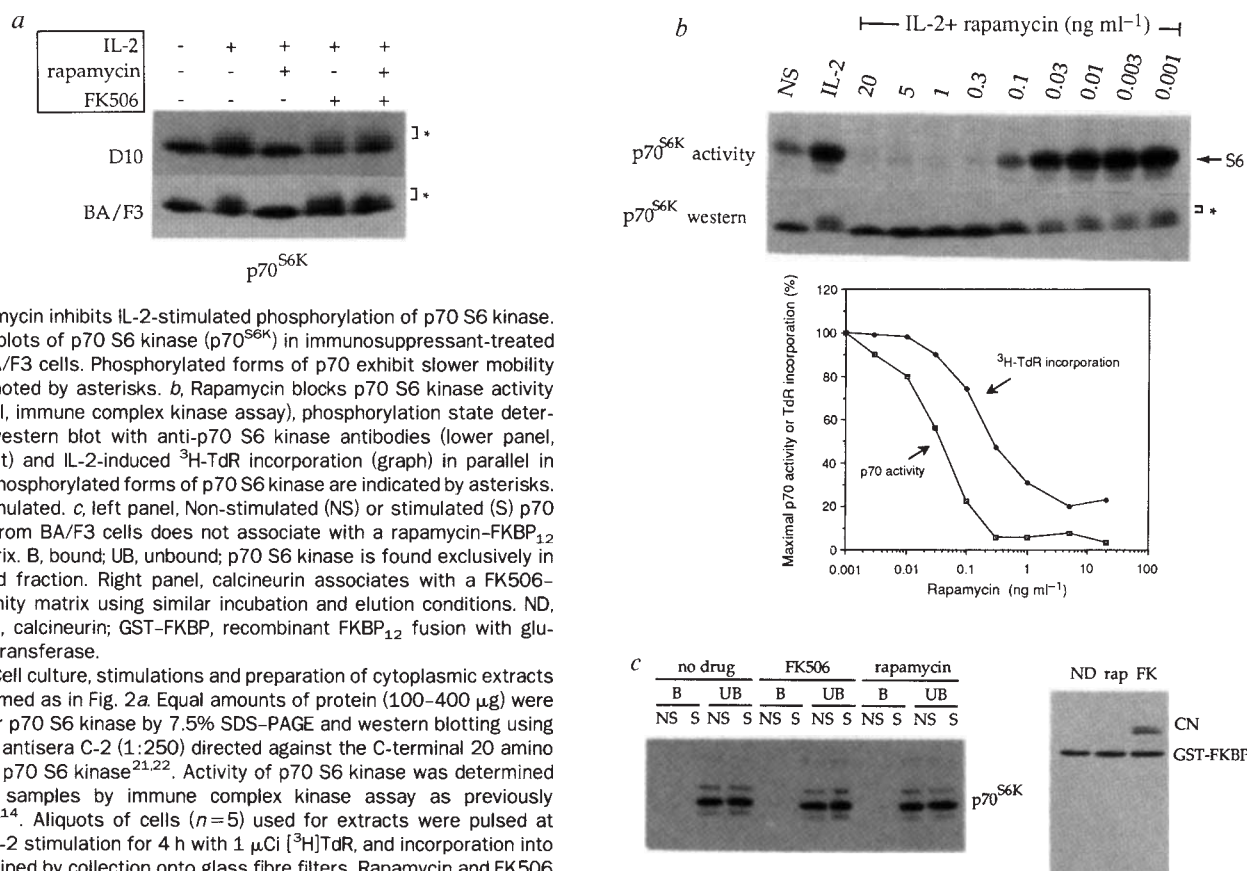


FIG. 3 Rapamycin inhibits IL-2-stimulated phosphorylation of p70 S6 kinase. **a**, Western blots of p70 S6 kinase (p70^{S6K}) in immunosuppressant-treated D10 and BA/F3 cells. Phosphorylated forms of p70 exhibit slower mobility and are denoted by asterisks. **b**, Rapamycin blocks p70 S6 kinase activity (upper panel, immune complex kinase assay), phosphorylation state determined by Western blot with anti-p70 S6 kinase antibodies (lower panel, Western blot) and IL-2-induced ³H-TdR incorporation (graph) in parallel in D10 cells. Phosphorylated forms of p70 S6 kinase are indicated by asterisks. NS, non-stimulated. **c**, left panel, Non-stimulated (NS) or stimulated (S) p70 S6 kinase from BA/F3 cells does not associate with a rapamycin-FKBP₁₂ affinity matrix. **B**, bound; **UB**, unbound; p70 S6 kinase is found exclusively in the unbound fraction. Right panel, calcineurin associates with a FK506-FKBP₁₂ affinity matrix using similar incubation and elution conditions. ND, no drug; CN, calcineurin; GST-FKBP, recombinant FKBP₁₂ fusion with glutathione-S-transferase.

METHODS. Cell culture, stimulations and preparation of cytoplasmic extracts were performed as in Fig. 2a. Equal amounts of protein (100–400 µg) were analysed for p70 S6 kinase by 7.5% SDS-PAGE and Western blotting using anti-p70^{S6K} antisera C-2 (1:250) directed against the C-terminal 20 amino acids of rat p70 S6 kinase^{21,22}. Activity of p70 S6 kinase was determined on parallel samples by immune complex kinase assay as previously described^{13,14}. Aliquots of cells ($n=5$) used for extracts were pulsed at 40 h post-IL-2 stimulation for 4 h with 1 µCi [³H]TdR, and incorporation into DNA determined by collection onto glass fibre filters. Rapamycin and FK506 affinity matrices were constructed as described¹⁹, containing 7.5 µg recombinant GST FKBP₁₂ (gift from P. N. Kao), 2 µM immunosuppressant, and glutathione-Sepharose (Pharmacia) as a 1:1 (v/v) slurry in 100 µl buffer A (ref. 19). Extract (250 µg) from IL-2-stimulated or non-stimulated BA/F3 cells were precleared with glutathione-sepharose (1 h, 4 °C, with rotation) in 100 µl buffer A (ref. 19) and then combined with 100 µl of the appropriate affinity matrix (2 h at 4 °C with rotation). The extract:matrix mixture was pelleted by centrifugation and the unbound material (UB) was collected as supernatant. Matrices were then washed extensively in buffer A, collected by centrifugation, and bound material (B) obtained by boiling the pellet in

SDS sample buffer. Total bound and unbound material was analysed by 8% SDS-PAGE and Western blotting for p70 S6 kinase using the C-2 antisera (1:250). Identical results were obtained when Ca²⁺ was omitted from buffer A (data not shown). In parallel, purified calcineurin (20 U; Sigma) and calmodulin (20 U; Sigma) were incubated with FKBP₁₂ affinity matrices containing no drug (ND), FK506 or rapamycin, as described¹⁹. Bound material was collected by boiling the pellets in SDS sample buffer, with subsequent separation by 8% SDS-PAGE and silver staining. Calmodulin also depended on Ca²⁺ and FK506/FKBP for retention but stained poorly.

with only a small proportion of cellular FKBP, in analogy with our proposal for the actions of FK506-FKBP (refs 4, 18).

The rapamycin block of IL-2-stimulated phosphorylation of p70 S6 kinase indicated that the rapamycin-FKBP complex did not directly inhibit p70 S6 kinase activity but rather interfered with signal transduction events upstream of p70 S6 kinase. To evaluate the competence of p70 S6 kinase to form ternary complexes with rapamycin-FKBP *vis-à-vis* the calcineurin-FK506-FKBP complex^{19,20}, we tested p70 S6 kinase binding to a rapamycin-FKBP₁₂ affinity matrix¹⁹. Under conditions in which calcineurin was retained on a FK506-FKBP₁₂ affinity matrix (Fig. 3c, right panel, and data not shown), phosphorylated or dephosphorylated p70 S6 kinase was not bound in

association with either rapamycin or FK506, in the presence (Fig. 3c) or absence (data not shown) of Ca²⁺; p70 S6 kinase was detected only in the unbound (UB) fraction. The inability of rapamycin affinity matrices to retain p70 S6 kinase is again consistent with a model in which rapamycin inhibits p70 S6 kinase activity through blockade of an upstream signal transduction pathway, although it is possible that our conditions are not permissive for the interaction of p70 S6 kinase with the rapamycin-FKBP complex. In this regard, rapamycin and recombinant FKBP do not affect p70 S6 kinase activity *in vitro*²³.

As rapamycin blocks T-cell entry into S phase^{2,4,5}, the selective, high-affinity and FKBP-dependent rapamycin antagonism of the p70 S6 kinase activation cascade (Figs 2, 3) implicates

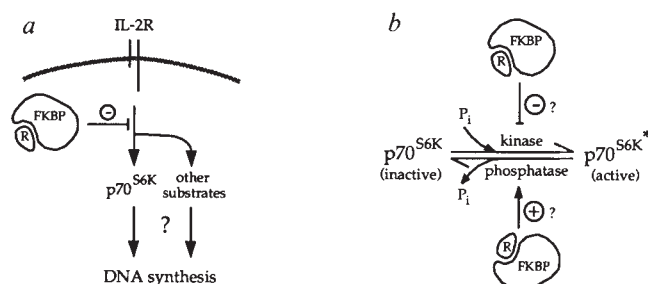


FIG. 4 **a**, A model for rapamycin blockade of IL-2-induced S phase entry. IL-2 selectively activates p70 S6 kinase but not MAP kinase or RSK, and this activation is antagonized by the rapamycin-FKBP complex. The rapamycin block of S phase entry may be mediated by inhibition of p70 S6 kinase (p70^{S6K}) or by inhibition of yet-undescribed molecules in the p70^{S6K} activation pathway. **R**, rapamycin. **b**, Potential mechanisms for rapamycin action on p70^{S6K}. The rapamycin-FKBP complex either inhibits a p70^{S6K} kinase or stimulates a p70^{S6K} phosphatase, driving the equilibrium phosphorylation state of p70 S6 kinase into the dephosphorylated, inactive form (leftwards). The action of the rapamycin-FKBP complex on p70-modifying enzymes may be direct or indirect.

molecules in this signalling pathway in the regulation of T lymphocyte G1 → S progression. We cannot currently distinguish between (1) direct p70 S6 kinase regulation of S phase entry or (2) branching of the p70 S6 kinase activation pathway to reach non-p70 S6 kinase substrates which themselves are crucial regulators of progression (Fig. 4a). The ability of rapamycin to block p70 S6 kinase activation by a wide variety of mitogens in several non-lymphoid cell types²³, suggests that the rapamycin target is not cell-type-specific and may assume a function central to numerous signal transduction pathways. The rapamycin-

induced accumulation of dephosphorylated p70 S6 kinase further suggests that FKBP-rapamycin may associate with and inhibit a p70 S6 kinase or stimulate a p70 S6 phosphatase, as outlined in Fig. 4b. Finally, the apparent reliance of the IL-2 receptor on the rapamycin-sensitive p70 S6 kinase signalling cascade (Fig. 1), as opposed to the more pleiotropic kinase activation engendered by other mitogens^{8,13,14}, could render IL-2-dependent cells uniquely susceptible to rapamycin, thereby explaining its selective immunosuppressive effects. □

Received 26 February; accepted 29 May 1992.

1. Heitman, J., Movva, N. R. & Hall, M. N. *Science* **253**, 905-909 (1991).
2. Morris, R. E. *Immun. Today* **5**, 137-140 (1991).
3. Dumont, F. J. et al. *J. Immun.* **144**, 1418-1424 (1990).
4. Bierer, B. E. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9231-9235 (1990).
5. Dumont, F. J., Staruch, M. J., Koprak, S. L., Melino, M. R. & Sigal, N. H. *J. Immun.* **144**, 251-258 (1990).
6. Schreiber, S. L. *Science* **251**, 283-287 (1991).
7. Erikson, R. L. *J. biol. Chem.* **266**, 6007-6010 (1992).
8. Blenis, J. *Cancer Cells* **3**, 1-4 (1991).
9. Siekierka, J. J., Hung, S. H. Y., Poe, M., Lin, C. S. & Sigal, N. H. *Nature* **341**, 755-757 (1989).
10. Harding, M. W., Galat, A., Uehling, D. E. & Schreiber, S. L. *Nature* **341**, 758-760 (1989).
11. Standaert, R. F., Galat, A., Uehling, D. E. & Schreiber, S. L. *Nature* **346**, 671-674 (1990).
12. Miyazaki, T., Maruyama, M., Yamada, G., Hatakeyama, M. & Taniguchi, T. *EMBO J.* **10**, 3191-3197 (1991).
13. Chung, J., Chen, R.-H. & Blenis, J. *Molec. cell Biol.* **11**, 1868-1874 (1991).
14. Chen, R.-H., Chung, J. & Blenis, J. *Molec. cell Biol.* **11**, 1861-1867 (1991).

15. Cantrell, D. A. & Smith, K. A. *Science* **224**, 1312-1316 (1984).
16. Pelech, S. L., Olwin, B. B. & Krebs, E. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5968-5972 (1986).
17. Price, D. J., Gunsalus, J. R. & Avruch, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7944-7948 (1990).
18. Schreiber, S. L. & Crabtree, G. R. *Immun. Today* **13**, 136-142 (1992).
19. Liu, J. et al. *Cell* **66**, 807-815 (1991).
20. Friedman, J. & Weissman, I. *Cell* **66**, 799-806 (1991).
21. Banerjee, P. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8550-8554 (1990).
22. Kozma, S. C. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7365-7369 (1990).
23. Chung, J., Kuo, C. J., Crabtree, G. R. & Blenis, J. *Cell* (in the press).

ACKNOWLEDGEMENTS. We thank R. Bram, N. Clipstone, R. L. Erikson, J. Jorgenson, B. Ho and C. Sagerstrom for discussion, J. Moore and S. Zurawski (DNAX) for cell lines and recombinant mouse IL-2, and P. Kao for GST-FKBP₁₂. This work was supported by the Howard Hughes Medical Institute (G.R.C.), the Lucille P. Markey Charitable Trust (J.B.), the American Cancer Society (J.B.) and the National Institutes of Health (G.R.C., J.B.). W.M.F. is an American Cancer Society Postdoctoral Fellow, and C.J.K. and D.F.F. were supported by the Medical Scientist Training Program at Stanford University School of Medicine.

Phosducin is a protein kinase A-regulated G-protein regulator

Petra H. Bauer*, Stefan Müller*†, Mechthild Puzicha*, Susanne Pippig*, Brigitte Obermaier*, Ernst J. M. Helmreich† & Martin J. Lohse*

* Laboratory of Molecular Biology, University of Munich, Max-Planck-Institute of Biochemistry, 8033 Martinsried, Germany
† Department of Physiological Chemistry, University of Würzburg, Medical School, 8700 Würzburg, Germany

SIGNAL transduction by G-protein-coupled receptors is regulated by various mechanisms acting at the receptor level; those studied most thoroughly are from the β -adrenergic receptor/ G_s /adenylyl cyclase system¹⁻³. We report here a regulatory mechanism occurring at the level of the G proteins themselves. A protein with M_r 33,000 that inhibits G_s -GTPase activity was purified from bovine brain. This protein is very similar or identical to phosducin, a protein previously thought to be specific for retina and pineal gland⁴⁻⁸. Recombinant phosducin inhibited the GTPase activity of several G proteins, and also inhibited G_s -mediated adenylyl cyclase activation. Blockade of its inhibitory effects by protein kinase A suggests that phosducin may be part of a complex regulatory network controlling G-protein-mediated signalling.

A major mechanism regulating β -adrenergic receptor function involves receptor phosphorylation catalysed by the β -adrenergic receptor kinase (β ARK), followed by binding of β -arrestin to the phosphorylated receptors⁹⁻¹². Crude preparations of β ARK from bovine brain cytosol already inhibit β_2 -adrenergic receptor function by up to 70% (ref. 10). This degree of inhibition suggests that, in addition to β -arrestin and β ARK, other inhibitors of the receptor- G_s interaction might be present. Indeed, when such preparations are chromatographed on DEAE-cellulose, fractions eluting at ~250 mM NaCl contain an additional inhibitory activity. In contrast to β -arrestin^{11,12}, this inhibitory activity is not dependent on previous receptor phosphorylation by β ARK (Fig. 1a). Purification of the inhibitor by several chromatographic steps ultimately yielded a protein with M_r 33,000 (33K) (Fig. 1b).

This protein inhibited not only the β_2 -receptor-stimulated GTPase activity of G_s , but also the intrinsic GTPase activity of

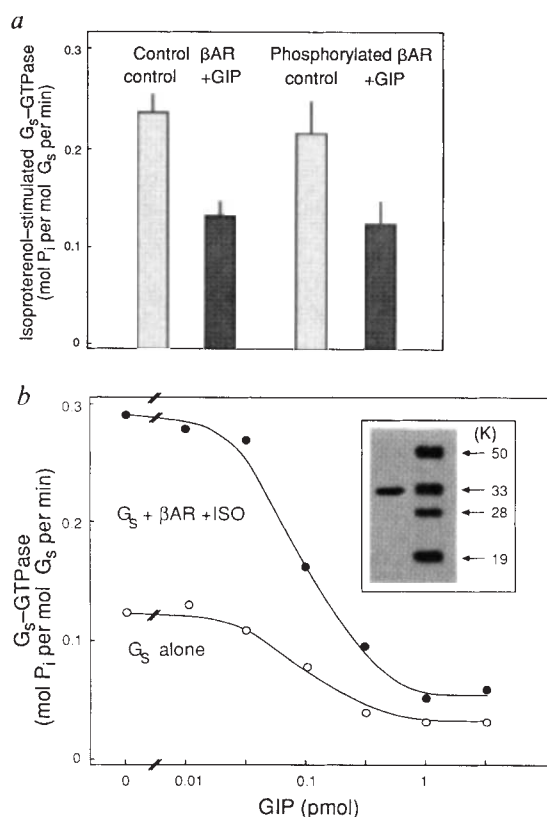
G_s alone (Fig. 1b). We therefore named it 'G-protein-inhibitor protein' (GIP). Half-maximal inhibition of the G_s GTPase activity occurred at a GIP: G_s ratio of ~2:1 (Fig. 1b). Assuming that the GIP preparation was neither completely homogeneous nor 100% active, we postulate that inhibition occurs by stoichiometric binding of GIP to G_s .

Microsequencing of CNBr-generated GIP fragments yielded two sequences: Ser-Ser-Pro-Gln-Ser-Arg-Asp-Asp-Lys and Ser-Val-Gln-Glu-Tyr-Glu-Leu-Ile. Both stretches are contained in a protein called phosducin⁷, the 33K protein⁷ or the MEKA protein⁶. Phosducin is thought to be specific for retina and pineal gland⁴⁻⁸, and has been purified from retina as a complex with the $\beta\gamma$ subunits of the retinal G protein, transducin⁴.

Phosducin messenger RNA was found in brain cortex by amplification with the polymerase chain reaction (PCR; Fig. 2). Identical PCR products were obtained with complementary DNA transcribed from retinal and from brain cortical mRNA. Their sequences were virtually identical to that of retinal phosducin and contained stretches coding for the two GIP fragments described above. Sequencing suggested further that there may be two variants of phosducin (Fig. 2 legend), which are both expressed not only in retina but also in brain cortex. The full-length PCR product from brain was overexpressed in *Escherichia coli*¹³ and purified to apparent homogeneity (Fig. 3 legend). The identity (or close similarity) of GIP and phosducin was further documented by the fact that antibodies against the synthetic peptide phosducin (97-111) recognized GIP and recombinant phosducin equally well (data not shown).

Recombinant phosducin inhibited the GTPase activity not only of G_s , but also of G_i and G_o equally well (Fig. 3a). Appreciable inhibition occurred at a phosducin: G -protein ratio of 2:1, but higher concentrations of phosducin were required than in the experiments shown in Fig. 1, suggesting a lower activity of the recombinant protein.

It has been suggested that retinal phosducin acts as a 'trap' for G protein $\beta\gamma$ subunits⁴. But phosducin also inhibited the GTPase activity of the $G\alpha_o$ subunit alone (Fig. 3b). In addition, phosducin reduced the stimulation of α_o GTPase activity which is caused by $\beta\gamma$ subunits, so that the $\beta\gamma$ subunits were required for the full inhibitory effects of phosducin. Thus, phosducin seems to interact both with the α_o -subunit and the $\beta\gamma$ subunits to form a complex of four proteins: phosducin and the α -, β - and γ -subunits of G_o .



Under basal conditions, the rate-limiting step of the G protein/GTPase cycle is the exchange of GDP for GTP¹⁴. When we measured this exchange as the rate of GTP- γ -³⁵S binding to G_0 , we observed no alteration of the rate constant by phosducin (Fig. 3c). At higher concentrations of phosducin, binding of GTP- γ -³⁵S was actually increased at all time points, suggesting a stabilization of the GTP-bound state of G_0 .

Phosphorylation of phosducin by protein kinase A (PKA) completely abolished its ability to inhibit G protein GTPase activity (Fig. 3d) but did not alter its stabilizing effect on GTP- γ -³⁵S binding (data not shown). These observations suggest that the regulatory role of phosducin in signal transmission may itself be subject to regulation by PKA.

In plasma membranes, phosducin inhibited signalling in the

FIG. 1 Identification and purification of a G-protein-inhibitor protein (GIP) from bovine brain. **a**, Inhibition of β_2 -adrenergic receptor (β AR)-stimulated G_s GTPase activity by crude GIP (eluting at ~ 250 mM NaCl from the DEAE-cellulose column) in a reconstituted phospholipid-vesicle system. Receptors were phosphorylated by β -adrenergic receptor kinase (β ARK) before the assays (right) or not (left). Means $\pm$ s.e.m. ($n=5$). **b**, Inhibition of G_s by purified GIP (inset) was added to vesicles containing either G_s alone, or G_s plus isoproterenol-activated β_2 -adrenergic receptors. Maximal inhibition (%) and potency (IC_{50} value) were calculated by nonlinear regression²⁰: G_s alone, 77% inhibition, IC_{50} at GIP: G_s 1.6; G_s + β_2 -receptors, 80% inhibition, IC_{50} at GIP: G_s 1.7 (means; $n=3$).

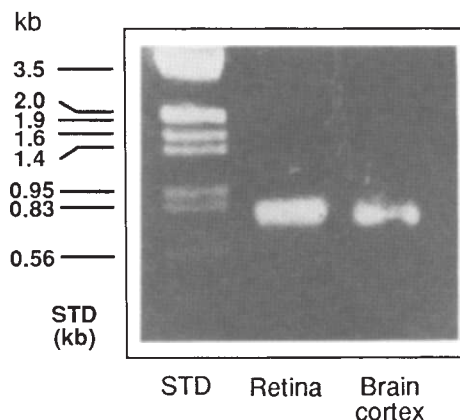
METHODS. A crude Ultrogel Aca34 β ARK-fraction prepared from bovine brain cortex⁹ was dialysed against 5 mM Tris-HCl, pH 8, and loaded onto a DEAE-Sephacel column, which was eluted with 0–500 mM NaCl in 5 mM Tris-HCl, pH 8. GTPase-inhibitory fractions (assayed as shown in **a**) eluting at ~ 250 mM NaCl were diluted threefold with 10 mM potassium phosphate buffer, pH 7.5 (PPB), and loaded onto an HA-Ultrogel (IBF) hydroxylapatite column, which was eluted with 10–300 mM PPB. Peak fractions eluting at about 100 mM PPB were diluted threefold and loaded onto a Mono-Q FPLC anion exchange column, eluted with 100–500 mM NaCl in 5 mM Tris-HCl, pH 7. Peak fractions eluting at 220–250 mM NaCl were concentrated and applied to a Superose-12 FPLC gel-filtration column. The inset shows a Coomassie-blue stain of the final peak fraction (right lane, protein standards). About 50 μ g GIP were obtained from 5 g cytosolic brain protein. Assuming a yield of $\sim 1\%$ in the six-step purification procedure, GIP may represent as much as 1/1,000 of the cytosolic brain protein. For GTPase assays, purified hamster lung β_2 -adrenergic receptors²¹ were reconstituted into phospholipid vesicles^{10,12}. They were phosphorylated or not by purified β ARK to >3 mol phosphate per mol receptor^{9,10} and purified G_s (ref. 22) was added. Tubes contained 100 fmol receptors (or none) and 50 fmol G_s in 100 μ l. Receptor-stimulated (in the presence of 10 μ M (–)-isoproterenol) and basal GTPase activities were determined¹² at 30 °C in the presence of 1 mg ml^{–1} BSA as ‘non-specific’ protein.

β_2 -adrenergic receptor/ G_s /adenylyl cyclase system when the system was stimulated at the receptor or G_s level, but not when it was activated at the cyclase level (Fig. 4), indicating an inhibitory effect of phosducin on G_s function. Thus, even though phosducin seems to stabilize G proteins in the GTP-bound state, this does not translate into an enhanced functional activity as might be anticipated. The postulated formation of a complex between phosducin and the trimeric G protein may inhibit G-protein-mediated signalling either by preventing the formation of an activated state after GTP-binding¹⁵, or by sterically interfering with G-protein-effector coupling.

The inhibition exerted by phosducin is of the same magnitude as that observed for rapid receptor-directed mechanisms, which may amount to up to 70–80% (refs 16, 17). Further experiments

FIG. 2 Identification of phosducin mRNA in bovine retina and brain by the polymerase chain reaction. The entire coding region of phosducin was amplified. Direct sequencing of the PCR products obtained from brain and retina revealed complete identity with each other, and were virtually identical to the published sequence of retinal phosducin^{5–7}. Differences are as follows: sequencing of PCR products repeatedly gave both A and C in nucleotide position 131, which makes codon 44 either CAC (His) or CCC (Pro). Subcloning of the PCR products resulted in sequences that contained unequivocally either CAC or CCC in this position. Both CAC (His)^{5,6} and CCC (Pro)⁷ have been reported for retinal phosducin, suggesting the existence of two variants. In position 84 we found a T in agreement with Lee *et al.*⁵ and Kuo *et al.*⁶, whereas Abe *et al.*⁷ found a C; both codons, GAT and GAC, code for Asp, however. The two protein sequences identified in GIP are contained in positions 54–62 and 73–80. STD, size standards in kilobases (kb).

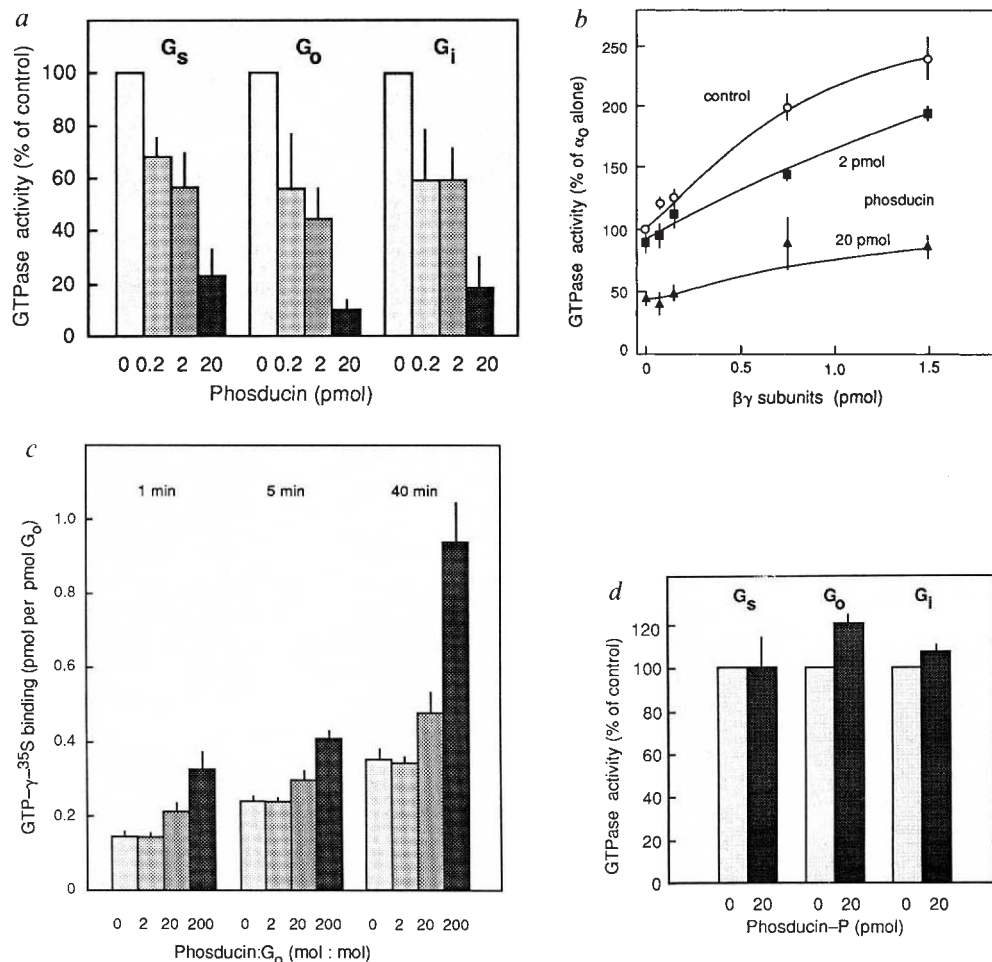
METHODS. Bovine brain cortex mRNA was prepared according to Chirgwin *et al.*²³, and 1 μ g was transcribed into cDNA in a volume of 50 μ l with MMLV reverse transcriptase (Superscript, BRL). Of this cDNA, 3 μ l were used as a template for PCR reactions using *Taq*-polymerase (Boehringer) according to the manufacturer's instructions. Primers (1 μ M) corresponding to the 5' and 3' ends of the phosducin cDNA^{5–7} were: 5'-ATGGAAAAAGCAAAAGCCAA, and 5'-TTATTCCATATCCTCTTCATGTT. After 3 min of denaturation at 94 °C, 29 cycles at 55 °C (1 min), 72 °C (1.5 min) 94 °C (1 min) plus a final cycle for 5 min at 72 °C were done. A sample (5 μ l) from the 100 μ l reaction mixture was run on a 1% agarose gel and the gel was stained with ethidium



bromide. The PCR products were isolated and sequenced directly on an automated system using *Taq* DNA polymerase and fluorescence-labelled dideoxynucleotides. Sequencing was done at least once in each direction and was confirmed by sequencing the PCR-product cloned into the *E. coli* expression vector pET8c (ref. 13) (see legend to Fig. 3).

FIG. 3 Effects of recombinant phosducin on G_s , G_o and G_i . *a*, Inhibition of the GTPase activity of G_s , G_o and G_i holoproteins by phosducin. *b*, Effects of phosducin on the GTPase activity of resolved α_o or $\alpha_o + \beta\gamma$ subunits. *c*, Stimulation of GTP- γ - 35 S binding to G_o by phosducin. *d*, Lack of inhibition of PKA-phosphorylated phosducin on the GTPase activity of G_o , G_s and G_i .

METHODS. Recombinant phosducin was purified from overexpressing *E. coli*. A phosducin-expression vector was constructed from pET8c¹³ using the *Nco*I and *Bam*HI sites in the vector and attaching the same sites to the phosducin PCR primers. The construct coded for His in position 44. BL21(DE3)pLysS *E. coli* were transformed, and induced for 2 h with 0.4 mM isopropylthiogalactoside¹³. DNA was precipitated from the freeze-thaw lysate with 2% streptomycin sulphate and centrifugation at 50,000g for 20 min. Phosducin was purified to apparent homogeneity from the supernatant by chromatography on Mono-Q and Superdex200 (last two steps, legend to Fig. 1). G_s and G_i were purified from rabbit liver^{22,23}, G_o and its resolved α - and $\beta\gamma$ subunits from bovine brain²⁵. GTPase assays¹² were done at 30 °C in solution using 0.1 pmol of holoproteins (*a*), or 0.15 pmol of α_o plus varying amounts of $\beta\gamma$ -subunits as indicated (*b*). Assay volumes were 100 μ l. *d*, Purified phosducin (20 μ g ml⁻¹) was phosphorylated before the GTPase assay with 1,000 U ml⁻¹ of the catalytic subunit of PKA²⁶. Phosphorylation of phosducin by PKA, assigned to Ser 73 (ref. 27), was verified by using γ -³²P-ATP, followed by electrophoresis and autoradiography (not shown). *a*, *d*, Activities of the G proteins alone were set to 100% and were 0.88 ± 0.22 (G_s), 1.53 ± 0.14 (G_o), and 1.36 ± 0.37 (G_i) mol P_i per 30 min per mol G protein, with the amount of functional G protein determined by equilibrium binding of GTP- γ - 35 S. *b*,



Values were normalized to the activity of α_o alone (2.1 ± 0.3 mol P_i per 30 min per mol α_o , determined as above). Binding of GTP- γ - 35 S was measured²⁸ with 1 μ M GTP- γ - 35 S at 15 °C. Phosducin itself did not bind GTP- γ - 35 S (not shown). All assays contained 1 mg ml⁻¹ BSA as 'non-specific' protein. Means $\pm$ s.e.m. ($n=4-7$).

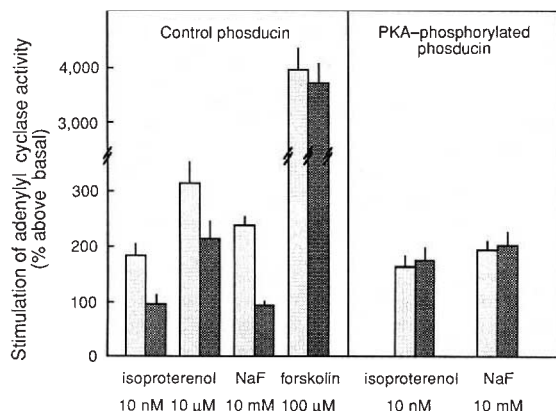


FIG. 4 Effects of phosducin on the β_2 -adrenergic receptor/ G_s /adenylyl cyclase system in A431 cell membranes. Left, non-phosphorylated phosducin; right, PKA-phosphorylated phosducin. Adenylyl cyclase activity was determined under basal conditions, with submaximal (10 nM) and maximal (10 μ M) stimulation of β_2 -adrenergic receptors by (–)isoproterenol, with stimulation of G_s by 10 mM NaF, and with stimulation of the cyclase itself by 100 μ M forskolin. Light bars indicate the activity in the absence, and darker bars the activity in the presence of phosducin. Means $\pm$ s.e.m. ($n=5$). Phosducin inhibited the isoproterenol- and NaF-stimulated activities but had no significant effects on forskolin-stimulated activity of adenylyl cyclase. The inhibition was most pronounced when the system was activated at the level of G_s with NaF (60% inhibition). In agreement with similar observations in the context of receptor desensitization²⁹, the inhibitory effects of phosducin were more pronounced at submaximal than at maximal stimulation of the β_2 -adrenergic receptors (44% and 30% inhibition at 10 nM and 10 μ M isoproterenol, respectively). When PKA-phosphorylated phosducin was used in such experiments, the inhibitory effects were completely abolished.

METHODS. Membranes were prepared from A431 cells and adenylyl cyclase activity was determined in these membranes as described earlier³⁰. Assays were done with 50 μ g of plasma membranes with or without 100 pmol phosducin in 100 μ l volume. The tubes also contained 100 μ g BSA. Basal activities in the absence and presence of phosducin were 8.9 ± 0.9 and 9.2 ± 1.1 pmol cAMP min⁻¹ mg⁻¹ membrane protein.

are required to determine whether cells contain enough phosphoducin to exert the full inhibitory effect. PKA-catalysed phosphorylation of phosphoducin abolished this inhibition (Fig. 4), suggesting that the net effect of phosphoducin in G_s -mediated signalling may be the relief from inhibition, that is, positive feedback. Conversely, if G_i does indeed mediate inhibition of PKA, this might increase its inhibitory potency, that is, cause negative feedback.

Expression of phosphoducin had been thought to be limited to retina and pineal gland⁴⁻⁸. But northern blots revealed the presence of its mRNA in brain, heart, kidney, liver, lung, spleen and skeletal muscle (data not shown). Its widespread distribution is reminiscent of other proteins involved in G-protein-mediated signalling. But whereas for all other functions, similar

but distinct proteins exist for visual and receptor-mediated signalling^{1,3}, we find apparently the same phosphoducin in retina and in brain cortex. Obviously, this does not exclude the existence of other similar proteins.

Recent observations have suggested that the growth-cone-associated protein GAP-43, and a complex between the small GTP-binding protein *ras* p21 and its GTPase-activating protein might be potential regulators of G-protein function^{18,19}. Phosphoducin may represent a particularly intriguing G-protein regulator, because its inhibitory activity is itself regulated by PKA. Its effects on multiple G proteins suggest that it may have a very general regulatory role in the complex network of mechanisms controlling transmembrane signalling systems. □

Received 15 January; accepted 15 May 1992.

1. Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. *A. Rev. Biochem.* **60**, 653-688 (1991).
2. Haddock, J. R. & Malbon, C. C. *Trends Neurosci.* **14**, 242-247 (1991).
3. Lohse, M. J. in *Signal Transduction in Photoreceptor Systems* (eds Hargrave, P. A., Hofmann, K. P. & Kaupp, U. B.) 160-171 (Springer, Berlin, 1992).
4. Lee, R. H., Lieberman, B. S. & Lolley, R. N. *Biochemistry* **26**, 3983-3990 (1987).
5. Lee, R. H., Flower, A., McGinnis, J. F., Lolley, R. N. & Craft, C. M. *J. Biol. Chem.* **265**, 15867-15873 (1990).
6. Kuo, C.-H. *et al. Biochem. biophys. Res. Commun.* **162**, 1063-1068 (1989).
7. Abe, T. *et al. Gene* **91**, 209-215 (1990).
8. Reig, J. A., Yu, L. & Klein, D. C. *J. Biol. Chem.* **265**, 5816-5824 (1990).
9. Benovic, J. L., Mayor, F., Staniszwski, C., Lefkowitz, R. J. & Caron, M. G. *J. Biol. Chem.* **262**, 9026-9032 (1987).
10. Benovic, J. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8879-8882 (1987).
11. Lohse, M. J., Benovic, J. L., Codina, J., Caron, M. G. & Lefkowitz, R. J. *Science* **248**, 1547-1550 (1990).
12. Lohse, M. J. *et al. J. Biol. Chem.* **267**, 8558-8564 (1992).
13. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185**, 60-89 (1990).
14. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
15. Birnbaumer, L. *A. Rev. Pharmac. Toxic.* **30**, 675-705 (1990).
16. Lohse, M. J., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *J. Biol. Chem.* **265**, 3202-3209 (1990).
17. Roth, N., Campbell, P. T., Caron, M. G., Lefkowitz, R. J. & Lohse, M. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6201-6204 (1991).

18. Yatani, A. *et al. Cell* **61**, 769-776 (1990).
19. Strittmatter, S. M., Valenzuela, D., Sudo, Y., Linder, M. E. & Fishman, M. C. *J. Biol. Chem.* **266**, 22465-22471 (1991).
20. Lohse, M. J., Klotz, K.-N. & Schwabe, U. *Molec. Pharmac.* **30**, 403-409 (1986).
21. Benovic, J. L., Shorr, R. G. L., Caron, M. G. & Lefkowitz, R. J. *Biochemistry* **23**, 4519-4525 (1984).
22. Sternweis, P. C., Northup, J. K., Smigel, M. D., Gilman, A. G. *J. Biol. Chem.* **256**, 11517-11526 (1981).
23. Chirwin, J. J., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
24. Bokoch, G. M., Katada, T., Northup, J. K., Uli, M. & Gilman, A. G. *J. Biol. Chem.* **259**, 3560-3567 (1984).
25. Sternweis, P. C. & Robishaw, J. D. *J. Biol. Chem.* **259**, 13806-13813 (1984).
26. Sugden, P. H., Holladay, L. A., Reimann, E. M. & Corbin, J. D. *Biochem. J.* **159**, 409-422 (1976).
27. Lee, R. H., Brown, B. M. & Lolley, R. N. *J. Biol. Chem.* **265**, 15860-15866 (1990).
28. Freissmuth, M. & Gilman, A. G. *J. Biol. Chem.* **264**, 21907-21914 (1989).
29. Lohse, M. J. *J. Biol. Chem.* **265**, 3210-3211 (1990).
30. Lohse, M. J., Lefkowitz, R. J., Caron, M. G. & Benovic, J. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3011-3105 (1989).

ACKNOWLEDGEMENTS. We thank C. Craft, M. Freissmuth and J. Elce for helpful discussions and comments at various stages of this work, G. Arnold for DNA and peptide synthesis, F. Hofmann for some batches of purified catalytic subunit of PKA and S. Andexinger for technical assistance. These studies were supported by the Bundesministerium für Forschung und Technologie and the Deutsche Forschungsgemeinschaft.

Antibody and HIV-1 gp120 recognition of CD4 undermines the concept of mimicry between antibodies and receptors

Simon J. Davis*, Gérard A. Schockmel*, Chamorro Somoza*, David W. Buck†, Don G. Healey‡, E. Peter Rieber§, Christian Reiter§ & Alan F. Williams*

* MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK

† BioTech Resources Inc., 12500 Network, Suite 112 San Antonio, Texas 78249, USA

‡ Department of Pathology, Cambridge University, Cambridge CB2 1QP, UK

§ Institut für Immunologie, Universität München, Goethestrasse 31, D-8000 München 2, Germany

IT has been proposed that antibodies can mimic the binding of a receptor to its ligand and that anti-idiotypic antibodies raised against such antibodies can be used to identify the receptor¹⁻³. A large number of antibodies have been raised against CD4, the receptor on T cells for the envelope glycoprotein gp120 of the human immunodeficiency virus, and the site at which gp120 binds to CD4 has been delineated⁴. It has therefore become possible to contrast the fine specificities of a natural ligand (gp120) and antibodies that interact with the receptor at the same site. Here we report that out of a panel of 225 anti-CD4 antibodies, only one showed fine binding specificity that was broadly like that of gp120, but the evidence was against this being an exact mimic. Thus the data indicate that the production of antibody mimics will occur very rarely or not at all and that the anti-idiotypic approach is unlikely to be useful. This contention is supported by a review

of the results of attempts to use this approach. Taking strict criteria for success, there is no example for which the anti-idiotypic approach has led to the discovery of a previously undescribed receptor or other protein of interest.

Mutants of rat CD4 have been constructed that bind gp120 with different affinities⁴. These mutants are based on constructs in which different residues, mostly in the region of residues 33-62 of domain 1 of rat CD4, were made identical to human CD4. It is argued that any anti-human CD4 antibody that does not bind rat CD4 but does react with a gp120-binding mutant must be binding at the gp120 binding site. Here we identify ten antibodies with this property and contrast the fine specificity of their binding with that of gp120.

The antibodies used here are summarized in Table 1. In Table 1a 98 antibodies are in the MT series, 31 of which bind to domain 1 of CD4. Among these 31 antibodies only nine inhibit gp120 binding and block syncytium formation at low concentrations. In Table 1b, the 11 antibodies were selected as possible gp120 mimics from 120 L-series antibodies⁵ on the basis that they bound to CD4 domain 1, inhibited gp120 binding and showed a pattern of loss of binding to mutated human CD4 similar to that of gp120 (refs 6,7). The seven antibodies shown in Table 1c had also been preselected on the basis that they inhibited gp120 binding.

The antibodies were initially tested for binding to COS-1 cells expressing the first rat CD4 mutant found to bind gp120 (ref. 4). This mutant, called mutant 2, has residues 36-62 identical to human CD4 and binds gp120 with an affinity 70-fold below that of human CD4 (ref. 4). Only 10 of the antibodies in Table 1, which were derived from an initial pool of 225 antibodies, bound to mutant 2. Representative data demonstrating the binding of the anti-human CD4 antibodies MT483 and MT4108 to mutant 2 is shown in Fig. 1a. It is noteworthy that the monoclonal antibody Leu3A, which has been proposed as a gp120 mimic^{8,9}, did not bind mutant 2. This result is consistent with recent reports showing that the immunization of humans and mice with Leu3A fails to induce anti-HIV activity¹⁰⁻¹².

TABLE 1 Production of antibodies and testing for binding to mutant 2 CD4

(a) Production of MT-series monoclonal antibodies (mAbs)

Total no. mAbs	Immunogen*	No. of fusions	No. of mAbs against domains†				Domain 1 mAbs judged to be potential gp120 mimics‡	mAbs binding mutant 2§
			1	1/2	2	3/4		
5	Human T cells	3	2	1	2	0	MT310	—
8	T-cell line + P815-CD4	4	1	6	0	1	—	—
23	P815-CD4	6	6	7	5	5	MT414, MT466	—
28	P815-CD4 + human sCD4	4	6	14	2	6	—	—
34	P815-CD4 + anti-CD4 mAb	4	16	3	12	3	MT483, MT494, MT495, MT497, MT4105, MT4108	MT4108, MT483
98 (total)		21	31	31	21	15	9	2

(b) L-series|| mAbs selected as possible mimics of gp120

Total no. mAbs	Potential mimics	mAbs binding mutant 2
120	Leu3A, L34, L80, L93, L197, L200, L201, L202, L222, L223, L252	L222, L223, L252

(c) Additional mAbs studied¶

Source	mAbs	mAbs binding mutant 2
AIDS-Directed Programme	6.2G4, 7.2G4, 7.3F11, 12.22.F5.C4, D4056	6.2G4, 7.2G4, 7.3F11
Others	30F16H5, NU-TH/1	30F16H5, NU-TH/1

* Immunizations: Human T cells: mice were immunized intraperitoneally with human thymocytes or T-cell lines, or cells from chronic lymphocytic T-cell leukaemias at least twice before fusion. P815-CD4: mice were immunized at least twice with P815 mouse mastocytoma transfectants expressing the complete human CD4 molecule. P815-CD4, human sCD4: mice were immunized at least twice with P815-CD4 transfectants. For the last boost 3 days before fusion, human sCD4 was used. The human sCD4 was prepared (gift from P. Tabaczewski and E. Weiss) by deleting the transmembrane region from amino acids 368–398, and this was secreted by CHO transfectants. For immunization, human sCD4 was bound to anti-CD4 mAbs specific for domains 3 and 4 which were bound to Sepharose-conjugated rabbit anti-mouse IgG. P815-CD4, CD4 mAb: mice were immunized several times intraperitoneally with P815-CD4 transfectants. Before fusion, P815-CD4 cells were incubated with four different CD4 mAbs binding to domains 3 and 4 (0.5 mg of each mAb) and injected together with the antibodies.

† Epitopes recognized by the different mAbs were localized on domains 1, 2 and 3 or 4 (3/4), or on conformational epitopes probably formed by domains 1 and 2 (1/2) by competition with reference mAbs previously characterized on CD4 mutants.

‡ Monoclonal antibodies were assigned as potential gp120 mimics when $3 \mu\text{g ml}^{-1}$ mAb was able completely to inhibit gp120 binding on CD4⁺ peripheral blood lymphocytes and when the same concentration completely blocked syncytia formation in 24 h co-cultures of HIV 1-infected H9 cells and uninfected C8166 cells.

§ The mutants were prepared as described in detail elsewhere⁴. Briefly, a rat CD4 cDNA was subcloned into M13mp19 for *in vitro* mutagenesis using a commercial kit (Amersham International, UK). The following sequences were introduced: mutant 1, Ile 36–Leu 51; mutant 2, Ile 36–Trp 62; mutant 6, Gln 33–Trp 62; mutant 7, Ile 36–Arg 59; mutant 8, Ile 36–Arg 58. Mutant 5 contained the human sequences Ile 36–Trp 62 and Ser 120–Ser 128, and lacked the second-domain glycosylation site. Mutant 9 contained the human sequences Ile 36–Trp 62 and the Gln residue at position 80 was mutated to the Asp present in human CD4. The mutants were then subcloned into the expression vector pCDM8 (refs 29, 30) or mutated using the polymerase chain reaction and inserted into the expression vector pEE6.hcmv-GS for expression in soluble form in Chinese hamster ovary cells as described¹³. Human and rat CD4 were expressed in the same ways for use as positive and negative controls. For immunofluorescence analysis of antibody binding, 4.5×10^6 COS-1 cells were transfected with the pCDM8 vectors containing the CD4 genes as described^{29,30}, and two days later the cells were collected and washed twice with phosphate-buffered saline (PBS), 0.2% bovine serum albumin (BSA). Cells (0.5×10^6) were incubated with tissue culture supernatant containing the appropriate anti-human CD4 antibody or the OX21 negative control mAb that recognizes human complement C3b inactivator. The cells were then left for 60 min at 4 °C, washed twice with PBS, 0.2% BSA, incubated with 40 μl fluorescein isothiocyanate-conjugated purified rabbit F(ab')₂ antimouse antibody at $10 \mu\text{g ml}^{-1}$ for 30 min at 4 °C and resuspended in PBS, 0.2% BSA for analysis on a FACScan flow cytometer (Becton Dickinson) using forward light scattering to gate out dead cells.

|| Leu3A and the L-series of mAbs were raised at Becton Dickinson. The mAbs were raised after immunization of BALB/c mice with the HPB-ALL cell line expressing high levels of human CD4.

¶ The mAbs 7.3F11, 6.3G4, 7.2G4 and 12.22.F5.C4 were obtained from the UK AIDS Directed Programme and were originally produced by J. Habeshaw and L. Walker; D4056 was made by D.H.; NU-TH/1, originally from the laboratory of M.M. Yokoyama, was obtained from the Third International Leukocyte Workshop panel; 30F16H5 was from the laboratory of F. Emmrich.

The fine specificities of the ten antibodies were compared with that of gp120 by determining their relative affinities for mutants 1, 2, 6, 7, 8 and 9 expressed in soluble form in CHO cells (a description of the mutants is given in the legend to Table 1; mutants are numbered as in ref. 4). The relative affinities were determined in inhibition assays with immobilized antibody or gp120 and with ¹²⁵I-labelled human soluble CD4 (sCD4) (Fig. 2). It has been shown elsewhere that CD4 does not dimerize at the concentrations used¹³, hence in this assay it is unlikely that there will be multivalent binding of the labelled ligand. The mutant forms were used to inhibit this binding assay with the results shown for all mutants except mutant 1 in Fig. 2. Mutant 1 does not bind gp120 or any of the antibodies, except 6.3G4 and 7.2G4, to which it bound with an affinity 200 times lower than human CD4 (data not shown). None of the antibodies had precisely the same binding pattern as gp120, yet the possibility that mimics could be detected using this assay was clearly

demonstrated by the finding that the following pairs of antibodies had exactly the same inhibition profiles: MT483 and MT4108, L222 and L223, 6.3G4 and 7.2G4. These pairs of antibodies also have the same kinetic properties (see later) and are therefore either antibodies with identical binding profiles raised independently or are derived from one original clone.

The only antibody that did not show major differences to gp120 with respect to its fine specificities for the mutants was the antibody 30F16H5 (Fig. 2). In this case, the relative inhibitory potency of the mutants was similar to that seen with gp120. But mutant 2 had a threefold higher relative affinity for 30F16H5 than for gp120, and mutant 8 had a slightly higher affinity for gp120 than mutant 2, whereas mutants 2 and 8 had the same affinities for 30F16H5. In addition, 30F16H5 binding was clearly distinguishable from that of gp120 in terms of binding to another rat CD4 mutant called mutant 5. Mutant 5 is based on mutant 2, with additional human sequences substituted in domain 2,

and it also has the rat CD4 glycosylation signal removed from this domain⁴. Antibody 30F16H5 binds mutant 5 expressed transiently on the surface of COS-1 cells, whereas gp120 binding is not detectable (Fig. 1b).

In addition to fine specificity, the antibodies differed from gp120 with respect to their binding kinetics. The kinetics were studied with immobilized gp120 or antibodies, and so the only factors determining the kinetics were the diffusion rate of the ¹²⁵I-labelled human sCD4 and the intrinsic properties of the binding sites. The association rate constant (k_{on}) values for the antibodies differed by only 3-fold, in contrast to a 35-fold variation in the dissociation rate constant (k_{off}) values (Table 2). Thus in this case the relative affinities of antibodies binding to the same region of a protein are largely determined by the k_{off} . However, wider divergence in on-rates can occur with antibodies binding to domain 1 of CD4 because the W3/25 antibody against rat CD4 has a k_{on} of $1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (ref. 14) and this was remeasured at $1.2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ by methods used here for the anti-human CD4 antibodies. This is three times faster than the fastest anti-human CD4 antibody that we measured. W3/25 binds to rat CD4 in an equivalent region to the human CD4 gp120 binding site because mutant 1 (described in legend to Table 1) does not bind W3/25 (data not shown). The pairs of antibodies with indistinguishable properties in the inhibition assays also have indistinguishable kinetic constants. The k_{on} for gp120 binding at 4 °C is 7–23-fold lower than that of the antibodies as a group, hence in this respect none of the antibodies is similar to gp120. Moreover, the antibodies are more similar as a group than they are similar to gp120. This implies that the interactions of antibodies and gp120 with CD4 may involve different types of binding sites or that regions outside the binding site affect the k_{on} . The k_{off} value for gp120 is also relatively low, but not outside the range observed for the antibodies.

We have established that less than five per cent of antibodies raised by immunization with CD4 bind to the same region as gp120 and that the specificities of the antibodies that do bind

in this region differ from those of gp120. This is consistent with the observation that anti-idiotypes raised against the antibodies MT483, MT4108, L252, 7.2G4, 7.3F11 and NU-TH/1 all fail to bind to gp120 (ref. 12, and D. H. unpublished results). This result is also in accord with another¹⁵ in which the three-dimensional structure of an anti-lysozyme antibody/anti-idiotypic complex was determined. The anti-idiotypic bound in a region that overlapped the binding site of lysozyme, yet the fine detail of the interactions was quite different. From the current data it seems likely that all of the anti-CD4 antibodies that bind the gp120 binding site will have differences in detail of binding to that of gp120 and it is not obvious why countless such variations should not occur.

An underlying concept behind the mimicry idea is that there will be a limited number of ways in which proteins might interact with a given region on another protein. For mimicry to occur the detail of the interaction has to be the same for a receptor and an antibody whose binding site is usually presented in a totally different structural context to the receptor's site. In contradiction to this is the long-standing tenet of immunochemistry that there are thousands of different ways in which an antibody binding site can be made against one determinant. The number has been estimated as 8,000 for the hapten dinitrophenol¹⁶, and the minimal repertoire of anti-haemagglutinin murine antibodies as 1,500 (ref. 17). By contrast, a dominant antibody response can occur at times in one mouse strain and it might be argued that this suggests a limited binding repertoire in some cases. But if the dominant antibody is suppressed then a new range of antibodies is seen¹⁸, again suggesting that the potential repertoire is extremely large. In the present

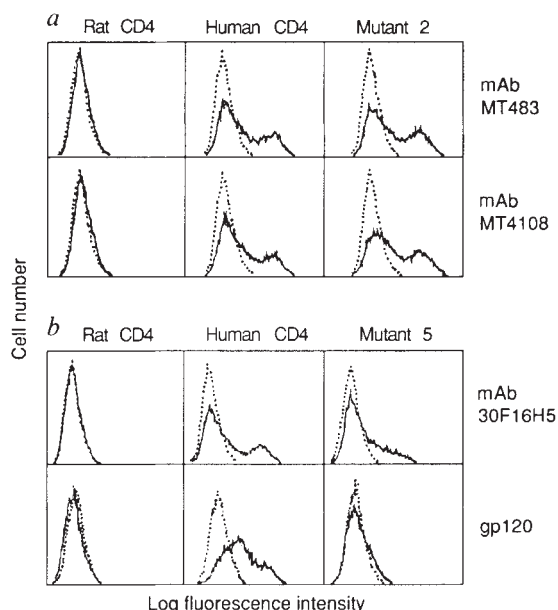


FIG. 1 a, Representative examples of the binding of antibodies to rat, human, and mutant 2 CD4 during the initial screening. b, Comparison of the specificities of gp120 and the mAb 30F16H5 for mutant 5 CD4. COS-1 cells were transiently transfected with pCDM8 (refs 29, 30) containing the genes for rat, human, mutant 2 and mutant 5 CD4, and then analysed by fluorescence-activated cell sorting after staining with antibodies MT483, MT4108 and 30F16H5, or with a combination of gp120 and the anti-gp120 mAb 108 (D.H., unpublished result) (heavy lines) as described in Table 1 legend. Dotted lines represent cells incubated with a negative control mAb, OX21.

TABLE 2 Kinetic parameters for the interactions of gp120 and mAbs with CD4

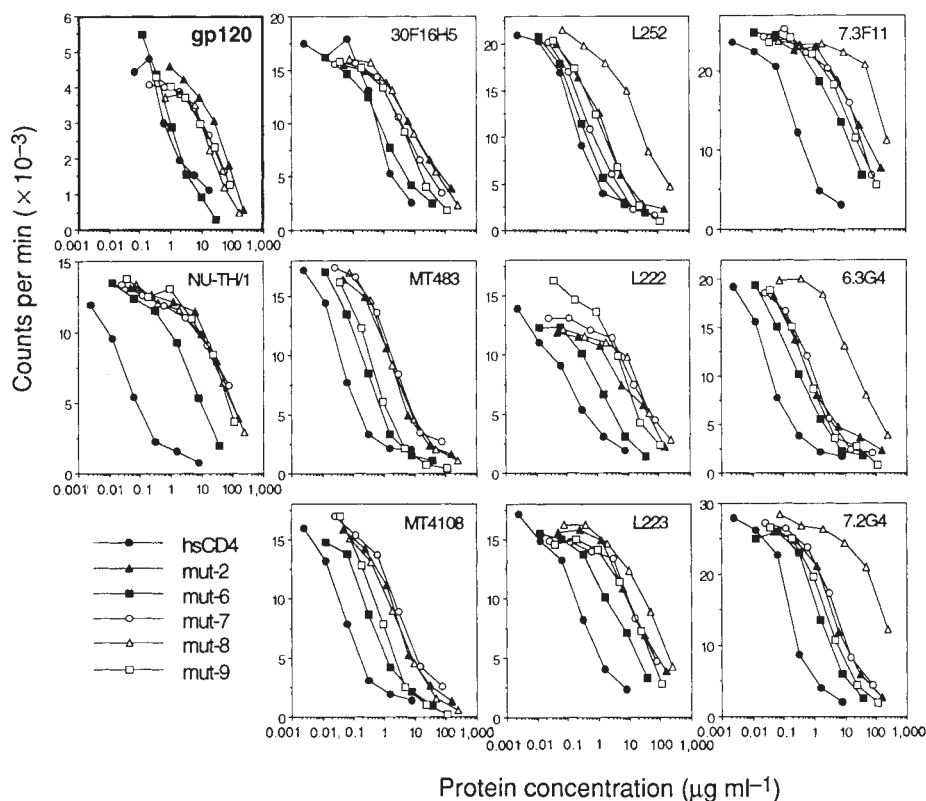
mAb/gp120	k_{on}^* $\text{M}^{-1} \text{ s}^{-1} \times 10^{-5}$	$k_{off}^\dagger$ $\text{s}^{-1} \times 10^5$	K $\text{M}^{-1} \times 10^{-9}$
MT483	2 ± 0.7	3.9 ± 0.8	5.1
MT4108	1.8 ± 0.6	4.2 ± 0.5	4.3
L222	3.5 ± 1.2	29 ± 0.58	1.2
L223	3.1 ± 0.8	29 ± 0.58	1.1
L252	1.1 ± 0.3	4.6 ± 0.9	2.4
6.3G4	2.2 ± 0.67	0.78 ± 0.2	28.0
7.2G4	2 ± 0.2	0.72 ± 0.2	28.0
7.3F11	1.3 ± 0.2	5.1 ± 0.49	2.5
30F16H5	1.1 ± 0.1	5.8 ± 0.7	1.9
NU-TH/1	3.1 ± 0.7	5.6 ± 0.42	5.5
gp120	0.15 ± 0.03	1.7 ± 0.5	0.88

* Dynabeads (Dyna) conjugated with sheep anti-mouse IgG were incubated with saturating amounts of the relevant antibody or with anti-gp120 mAb 108 and then gp120. The site numbers for the beads were determined by incubating the beads overnight with known saturating amounts of human sCD4 and trace amounts of ¹²⁵I-labelled human sCD4 (prepared as described in Fig. 2 legend) and by determining the proportion of label bound. The maximum ¹²⁵I-labelled human sCD4 that could be bound by antibody was routinely 80–90%. The k_{on} was determined essentially as described in ref. 14. The respective Dynabead targets or Dynabeads conjugated with a control antibody (MRC OX-7) were incubated with 800–1,000 ng ml⁻¹ ¹²⁵I-labelled human sCD4 in 2% BSA/PBS; the actual number of beads used was adjusted in accordance with the magnitude of the k_{on} . After different intervals 25-μl aliquots of the beads were removed and added directly to 1 ml 2% BSA/PBS in a microfuge tube. Beads were pelleted for 5–10 s and then washed again. Radioactivity associated with the pellets was determined by gamma counting. The k_{on} was calculated as described¹⁴.

† For the k_{off} determinations, Falcon 3911 96-well plates coated with rabbit anti-mouse IgG and the relevant antibody, or anti-gp120 mAb 108 and gp120, were prepared as described in Fig. 2 legend. These were then incubated with 50–200,000 c.p.m. of the ¹²⁵I-labelled human sCD4 in 25 μl 2% BSA/PBS overnight at 4 °C. At various times a 250-fold excess of unlabelled human sCD4 in BSA/PBS was then added to the wells. At the end of the incubation the plate was washed and the radioactivity bound was determined by gamma counting. k_{off} was determined as described¹⁴.

FIG. 2 Relative affinities of human CD4 and human-rat chimaeric CD4 molecules (mutants 2, 6, 7, 8, 9) for gp120 and anti-human CD4 antibodies.

METHODS. Mutants were prepared and expressed as described in Table 1 legend. For the gp120-inhibition assays, the wells of Falcon 3911 96-well plates were coated with rabbit anti-mouse IgG at $50 \mu\text{g ml}^{-1}$ for 45 min at room temperature and then washed and incubated with tissue culture supernatant containing the anti-gp120 mAb 108 overnight at 4°C . Plates were then washed and incubated with $100 \mu\text{l}$ recombinant soluble gp120 at 300 ng ml^{-1} in tissue culture supernatant (MRC AIDS Directed programme) for 5 h at 4°C . After a final wash the plates were incubated with $20\text{--}30 \mu\text{l}$ of $300\text{--}320 \text{ ng ml}^{-1}$ human sCD4 labelled to specific activity of $10\text{--}15 \mu\text{Ci } \mu\text{g}^{-1}$ with ^{125}I using the Bolton-Hunter reagent³¹ (Amersham), together with serial dilutions of unlabelled purified human sCD4 or 5-fold-concentrated tissue culture supernatants containing the soluble human-rat chimaeric proteins at different known concentrations. The antibody inhibition assays were performed in the same manner using antibody at 200 ng ml^{-1} or at $1/150\text{--}300,000$ dilutions of ascites in the place of mAb 108 and gp120 and $75\text{--}100 \text{ ng ml}^{-1}$ ^{125}I -labelled human sCD4 with the various inhibitors. After overnight incubation the plates were washed and the radioactivity bound to the plates determined by gamma counting. The data shown are representative of 4–8 determinations.



case one first has to find an antibody that binds to the relevant site in the target protein, and for CD4 the best frequency we can give is 5/218 antibodies for the MT- and L-series, for which the number of primary antibodies is known. The frequency of mimics is $<1/225$ for all the antibodies examined. The same problems of improbability apply again in the second round of immunization to produce the anti-idiotypic antibody.

Despite the problems illustrated by these probabilities, many workers may argue that the anti-idiotypic approach has been shown to work in model systems, going back to the first apparent success in 1978 (ref. 19). A more stringent test is to ask how many novel molecules or receptors have been discovered by this approach. We suggest that the criteria for success with the idiotypic approach should be that a molecule is discovered, the complementary DNA clone isolated, and the function proved by conferring the relevant activity on a negative cell line by transfection with the cDNA. According to these criteria we know of no accepted discovery of a receptor or novel molecule of proven importance. Discovery of the reovirus receptor as a molecule related to the adrenergic receptor was claimed²⁰, but

this has not been substantiated by cDNA cloning and transfection to yield a cell that binds the reovirus. Moreover, there are data that directly argue against the adrenergic receptor functioning as the reovirus receptor²¹ and sialic acid residues may mediate reovirus attachment^{22–24}. The identification of a protein import receptor for chloroplasts was claimed²⁵, but this molecule was then shown to be the phosphate translocator²⁶. There must also be doubts concerning the yeast mitochondrial protein import receptor proposed on the basis of anti-idiotypic data²⁷. A candidate for the KDEL receptor of endoplasmic reticulum was identified² but this remains to be substantiated; an HDEL receptor that is not the homologue of the molecule found by the anti-idiotypic approach has been cloned from yeast and a mammalian homologue identified²⁸. We argue that the underlying concept behind the mimicry idea is inconsistent with what is known about protein recognition and that the anti-idiotypic approach to receptor discovery has been found wanting in terms of yielding novel and important findings. Further applications seem equally unlikely to be rewarding. □

Received 2 April; accepted 15 May 1992.

- Gaulton, G. N. & Greene, M. I. *Rev. Immun.* **4**, 253–280 (1986).
- Vaux, D., Tooze, J. & Fuller, S. *Nature* **345**, 495–502 (1990).
- Vaux, D. J. T., Helenius, A. & Mellman, I. *Nature* **336**, 36–42 (1988).
- Schockmel, G. A., Somoza, C., Davis, S. J., Williams, A. F. & Healey, D. *J. exp. Med.* **175**, 301–304 (1992).
- Estess, P., Salmon, S. L., Winberg, M., Oi, V. T. & Buck, D. W. in *Current Research in Protein Chemistry* (ed. Villafranca, J.) 499–508 (Academic, New York, 1990).
- Sattentau, Q. J., Dalgleish, A. G., Weiss, R. A. & Beverley, P. C. L. *Science* **234**, 1120–1123 (1986).
- Sattentau, Q. J. *et al. J. exp. Med.* **170**, 1319–1334 (1989).
- Chan, T. C., Dreesman, G. R. & Kennedy, R. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3891–3895 (1987).
- Dalgleish, A. G., Thomson, B. J., Chan, T. C., Malkovsky, M. & Kennedy, R. C. *Lancet* **2**, 1047–1050 (1987).
- Wilks, D., Walker, L. C., Habeshaw, J. A., Webster, A. D. B. & Dalgleish, A. G. *J. infect. Dis.* **163**, 389–392 (1991).
- Reeves, J. P., Buck, D., Berkower, I., Murphy, D. & Epstein, S. L. *AIDS Res. hum. Retrovir.* **7**, 55–63 (1991).
- Healey, D. G. *et al. Eur. J. Immun.* **21**, 1491–1498 (1991).
- Davis, S. J. *et al. J. biol. Chem.* **265**, 10410–10418 (1990).
- Mason, D. W. & Williams, A. F. *Biochem. J.* **187**, 1–20 (1980).
- Bentley, G. A., Boulton, G., Rottot, M. M. & Poljak, R. J. *Nature* **348**, 254–257 (1990).

- Kreth, H. W. & Williamson, A. R. *Eur. J. Immun.* **3**, 141–147 (1973).
- Staudt, L. M. & Gerhard, W. *J. exp. Med.* **157**, 687–704 (1983).
- Eichmann, K. *Eur. J. Immun.* **4**, 296–302 (1974).
- Sege, K. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2443–2447 (1978).
- Co, M. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5315–5318 (1985).
- Choi, A. H. C. & Lee, P. W. K. *Virology* **163**, 191–197 (1988).
- Gentsch, J. R. & Pacitti, A. F. *Virology* **161**, 245–248 (1987).
- Paul, R. W. & Lee, P. W. K. *Virology* **159**, 94–101 (1987).
- Paul, R. W., Choi, A. H. C. & Lee, P. W. K. *Virology* **172**, 382–385 (1989).
- Pain, D., Kanwar, Y. S. & Blobel, G. *Nature* **331**, 232–237 (1988).
- Flügge, U.-I. *et al. Nature* **353**, 364–367 (1991).
- Pain, D., Murakami, H. & Blobel, G. *Nature* **347**, 444–449 (1990).
- Lewis, M. J. & Pelham, H. R. B. *Nature* **348**, 162–163 (1990).
- Seed, B. *Nature* **329**, 840–842 (1987).
- Seed, B. & Aruffo, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3365–3369 (1987).
- Bolton, A. E. & Hunter, W. M. *Biochem. J.* **133**, 529–539 (1973).

ACKNOWLEDGEMENTS. We are grateful for reagents supplied by the Medical Research Council AIDS Directed Programme. We also thank D. Mason for advice on kinetics. S.J.D. and C.S. are funded by the Medical Research Council AIDS Directed Programme; G.A.S. is supported by the Ministry of Cultural Affairs, Luxembourg; E.P.R. and C.R. are funded by the AIDS-directed FVP 1 of the BMFT and by the Deutsche Forschungsgemeinschaft.

Amplification of a gene encoding a p53-associated protein in human sarcomas

J. D. Oliner*, K. W. Kinzler*†, P. S. Meltzer‡, D. L. George§ & B. Vogelstein*

* The Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, Maryland 21231, USA

‡ Departments of Pediatrics and Radiation Oncology, University of Michigan Cancer Center, MSRB 11 C560B, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109, USA

§ Department of Human Genetics, University of Pennsylvania, Clinical Research Building, Philadelphia, Pennsylvania 19104, USA

DESPITE extensive data linking mutations in the p53 gene to human tumorigenesis¹, little is known about the cellular regulators and mediators of p53 function. MDM2 is a strong candidate for one such cellular protein; the MDM2 gene was originally identified by virtue of its amplification in a spontaneously transformed derivative of mouse BALB/c cells² and the MDM2 protein subsequently shown to bind to p53 in rat cells transfected with p53 genes^{3,4}. To determine whether MDM2 plays a role in human cancer, we have cloned the human MDM2 gene. Here we show that recombinant-derived human MDM2 protein binds human p53 *in vitro*, and we use MDM2 clones to localize the human MDM2 gene to chromosome 12q13–14. Because this chromosomal position appears to be altered in many sarcomas^{5–7}, we looked for changes in human MDM2 in such cancers. The gene was amplified in over a third of 47 sarcomas, including common bone and soft tissue forms. These results are consistent with the hypothesis that MDM2 binds to p53, and that amplification of MDM2 in sarcomas leads to escape from p53-regulated growth control. This mechanism of tumorigenesis parallels that for virally-induced tumours^{8,9}, in which viral oncogene products bind to and functionally inactivate p53.

To obtain human complementary DNA clones, a murine MDM2 cDNA probe was used to initiate cDNA walking in a human library (see legend to Fig. 1). Sequence analysis of 25 clones revealed several cDNA forms indicative of alternative splicing. The predominant human form is compared with its murine counterpart in Fig. 1. There was an open reading frame extending from the 5' end of the human cDNA sequence to nucleotide 1,784. Although this signal for translation initiation could not be unambiguously defined, the ATG at nucleotide 312 was considered the most likely position for several reasons. First, the sequence similarity between human and mouse MDM2 declined dramatically upstream of nucleotide 312. Second, an inverse polymerase chain reaction (PCR) was used in an attempt to acquire additional upstream cDNA sequence¹⁰. The 5' ends of the PCR-derived clones were very similar (within 12 base pairs) to the 5' ends of clones obtained from the cDNA library, indicating that the 5' end of the human MDM2 sequence shown in Fig. 1 may represent the 5' end of the transcript. Third, *in vitro* translation of the sequence shown in Fig. 1, beginning with the methionine encoded by the ATG at position 312, generated a protein similar in size to that observed in human cells (see below).

Comparison of the human and mouse MDM2 coding regions showed that they were 80.3% identical and shared a basic nuclear localization signal at codons 181 to 185 (ref. 11), several casein kinase II serine-phosphorylation sites¹², an acidic activation domain at codons 223 to 274 (ref. 13), and two metal-binding sites at codons 305 to 322 and 461 to 478, neither of which is highly related to known DNA-binding domains¹⁴.

To determine whether the human MDM2 protein could bind to human p53 protein *in vitro*, a human MDM2 expression vector was constructed from the cDNA clones (see legend to Fig. 2). RNA transcribed from this vector using T7 RNA polymerase was used to program a rabbit reticulocyte lysate. Although the predicted size of the protein generated from the construct was only 55.2K (M_r 55,200, extending from the methionine at nucleotide 312 to nucleotide 1,784), protein translated *in vitro* migrated at ~90 K. The MDM2 protein was not immunoprecipitated with antibodies against either the C-terminal or N-terminal regions of p53 (Fig. 2, lanes 2 and 3). But when *in vitro*-translated human p53 was mixed with the human MDM2 translation product, the anti-p53 antibodies precipitated MDM2 protein with p53 (Fig. 2, lanes 5 and 6). As a control, a protein of similar electrophoretic mobility (MCC¹⁵) was mixed with p53 and there was no coprecipitation (Fig. 2, lanes 8 and 9). When an *in vitro*-translated His-175 mutant form of p53 was mixed with human MDM2 protein, a similar coprecipitation of MDM2 and p53 proteins was also observed (data not shown).

Polyclonal rabbit antibodies were raised against an *Escherichia coli*-produced human MDM2–glutathione S-transferase fusion protein. The anti-MDM2 antibodies immunoprecipitated p53 when mixed with MDM2 protein (Fig. 2, lane 15) but failed to precipitate p53 alone (Fig. 2, lane 13).

To establish the chromosomal localization of human MDM2, somatic cell hybrids were screened, and a human–hamster hybrid containing only human chromosome 12 hybridized to the human MDM2 probe. Screening of hybrids containing portions of chromosome 12 (ref. 16) with the same probe narrowed the localization to chromosome 12q13–14. Because this region of chromosome 12 is often aberrant in human sarcomas^{5–7}, southern blot analysis to evaluate whether MDM2 was genetically altered in such cancers. We found a striking amplification of MDM2 sequences in several of these tumours (see examples in Fig. 3, lanes 2, 3 and 5). Of 47 sarcomas analysed, 17 showed a 5–50-fold MDM2 amplification. These tumours included 7 of 13 liposarcomas, 7 of 22 malignant fibrous histiocytomas, 3 of 11 osteosarcomas, and 0 of 1 rhabdomyosarcoma. Five benign soft tissue tumours (lipomas) and seventy-four carcinomas (colorectal or gastric) were also analysed by Southern blotting and no amplification was seen.

We next determined whether this gene amplification was associated with increased expression. Because of RNA degradation in primary sarcomas, only the cell lines could be productively analysed by northern blotting. In the one available sarcoma cell line with MDM2 amplification, a single transcript of ~5.5 kilobases (kb) was observed (Fig. 4a, lane 1). The amount of this transcript was much higher than in a sarcoma cell line without amplification (Fig. 4a, lane 2) or in a carcinoma cell line (Fig. 4a, lane 3). When purified messenger RNA (rather than total RNA) from the carcinoma cell line was used for analysis, a human MDM2 transcript of 5.5 kb could also be observed (Fig. 4a, lane 4). Expression of the MDM2 RNA in the sarcoma with amplification was estimated to be at least 30-fold higher than that in the other lines examined. This was consistent with results from western blot analysis. A protein of M_r ~90K was expressed at high levels in the sarcoma cell line with MDM2 amplification (Fig. 4b, lane 3), whereas no expression was evident in two sarcoma cell lines without amplification or in the carcinoma cell line (Fig. 4b, lanes 1, 2 and 4). Five primary sarcomas were also analysed by western blotting. Three primary sarcomas with amplification expressed the same sized protein as that in the sarcoma cell line (Fig. 4c, lanes 1–3), but no protein was observed in the two sarcomas without amplification (Fig. 4c, lanes 4 and 5).

Our results demonstrate that human MDM2 binds to p53 *in vitro* and is genetically altered in a significant fraction of the most common sarcomas of soft tissue and bone^{17,18}. It is important to note, however, that amplifications in human tumours

† To whom correspondence should be addressed.

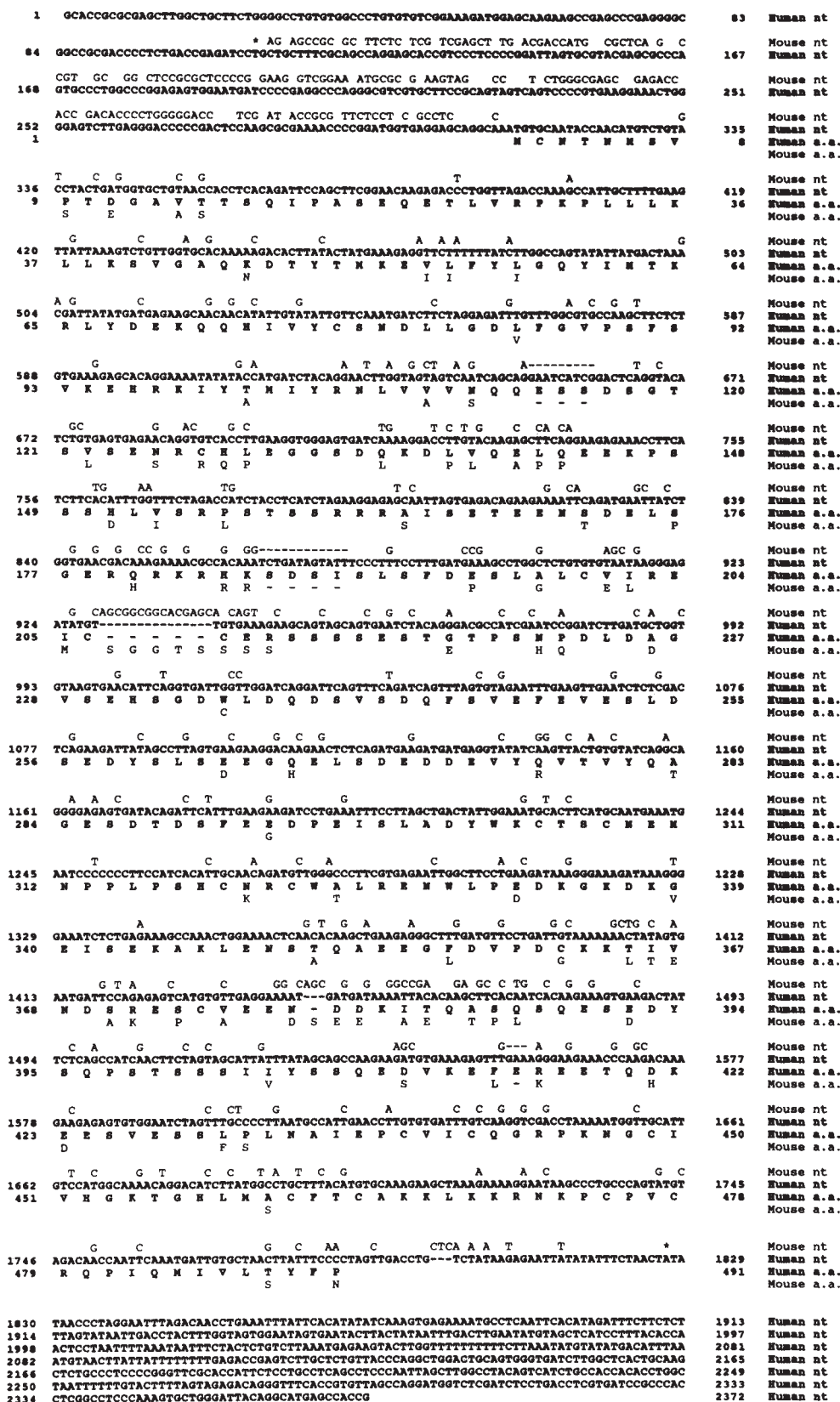


FIG. 1 Human *MDM2* cDNA sequence: human and mouse *MDM2* nucleotide (nt) and amino-acid (a.a.) sequences are compared (single-letter code). The mouse sequence is only shown where it differs from the corresponding human sequence. Asterisks mark the 5' and 3' boundaries of the previously published mouse *MDM2* cDNA². Dashes indicate insertions.

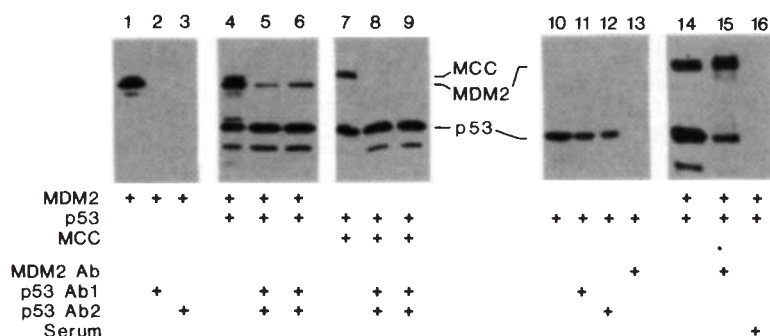
METHODS. Polyadenylated RNA isolated from the human colonic carcinoma cell line CaCo-2 was used as a template for the production of random hexamer-primed double-stranded cDNA²². The cDNA was ligated to adaptors and then to the lambda YES phage vector, packaged and plated as described²³. The library was screened initially with a ³²P-labelled ²⁴ mouse *MDM2* cDNA probe (nt 259 to 1,508; ref. 2) and then rescreened with a human *MDM2* cDNA clone containing nt 40 to 702. In total, 25 clones were obtained, partially or totally sequenced, and mapped. The sequence shown is representative of the most abundant class and was assembled from three clones: c14-2 (nt 1-949), c89 (nt 467-1,737), and c33 (nt 390-2,372). The 3' end of the untranslated region has not yet been cloned. The 5' end is likely to be at or near nt 1 (see text). The mouse and human amino-acid sequences are compared from the putative translation start site at nt 312 to the conserved stop codon at nt 1,784. This nucleotide sequence has been deposited at the EMBL Data Library, accession number Z12020.

often involve large stretches of the genome, encompassing 300 to 1,000 kb^{7,19-21}. Other genes in the *MDM2* amplicon could contribute to, or be responsible for, the growth advantage afforded by the amplification event. Nevertheless, *MDM2* is a good candidate for the 'target' of amplification for two reasons. First, *MDM2* has oncogenic activity after transfection into NIH3T3 cells². Second, it binds to a protein (p53) with known

growth suppressive effects on a wide variety of human tumour types. MDM2 may functionally inactivate p53 in ways similar to those employed by virally encoded oncoproteins such as simian virus 40 T antigen, adenovirus E1B and human papilloma virus E6 (refs. 4, 8, 9). Consistent with this hypothesis, no sarcomas with *MDM2* amplification (of five tested) had any of the p53 gene mutations that occur commonly in other tumours

FIG. 2 Coprecipitation of human MDM2 and p53. *In vitro*-translated MDM2, p53 and MCC proteins were mixed as indicated and incubated with p53 Ab1 (monoclonal antibody specific for the C terminus of p53), p53 Ab2 (monoclonal antibody specific for the N terminus of p53), MDM2 Ab (polyclonal rabbit anti-human MDM2 antibodies), or serum (preimmune serum obtained from the rabbit that produced the MDM2 antibody). Lanes 1, 4, 7, 10 and 14 contain aliquots of the protein mixtures used for immunoprecipitation. Bands running slightly faster than p53 are polypeptides produced from internal translation initiation sites.

METHODS. A human *MDM2* expression vector was constructed in pBluescript SK+ (Stratagene) from overlapping cDNA clones. The construct contained the sequence shown in Fig. 1 from nt 312 to 2,176. A 42-bp black beetle virus ribosome entry sequence²⁵ was placed immediately upstream of this *MDM2* sequence in order to obtain high expression. This construct, as well as p53 (ref. 26) and MCC¹⁵ constructs in pBluescript SK+, were transcribed with T7 RNA polymerase and translated in a rabbit reticulocyte lysate (Promega) according to the manufacturer's instructions. Lysate (10 μ l) containing the three proteins, alone or mixed in pairs, was incubated at 37 °C for 15 min. 1 μ g (10 μ l) of p53 Ab1 or Ab2 (Oncogene Science) or 5 μ l of rabbit serum containing MDM2 antibody or preimmune rabbit serum, were added as indicated. 90 μ l RIPA buffer (10 mM Tris, pH 7.5, 1% sodium deoxycholate, 1% NP40, 150 mM NaCl, 0.1% SDS), SNNT buffer³, or binding buffer²⁶ were then added and the mixtures allowed to incubate at 4 °C for 2 h. The three buffers produced similar results, although the



coprecipitation was less efficient in SNNT buffer (containing 0.5 M NaCl; lanes 5 and 8) than in binding buffer (containing 0.1 M NaCl; lanes 6 and 9). Following addition of 2 mg protein A-Sepharose, the tubes were rotated end-over-end at 4 °C for 1 h. After pelleting and washing, immunoprecipitates were electrophoresed on SDS-polyacrylamide gels and the dried gels autoradiographed in the presence of Enhance (New England Nuclear). Rabbits were immunized with a glutathione S-transferase (Pharmacia)-MDM2 fusion protein containing human MDM2 from the region corresponding to nt 390–816.

FIG. 3 Amplification of the human *MDM2* gene in sarcomas. DNA (5 μ g) was digested with *Eco*RI, separated by agarose gel electrophoresis and transferred to nylon as described²⁷. Filters were then hybridized with a human *MDM2* cDNA fragment probe (nt 1–949; Fig. 1) or to a control probe that identifies fragments of similar size (pDCC 1.65; ref. 28). Hybridization was as previously described²⁹. DNA was derived from 5 primary sarcomas (lanes 1–4, 6) and one sarcoma cell line (OsA-CL, lane 5). On longer exposure, the same sized *MDM2* fragments were observed in lanes 1, 4 and 6. DNA fragment sizes are shown on the left in kb.

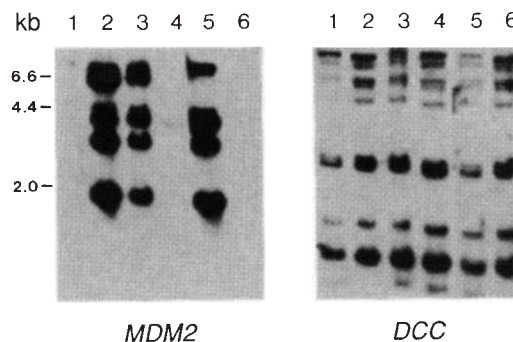
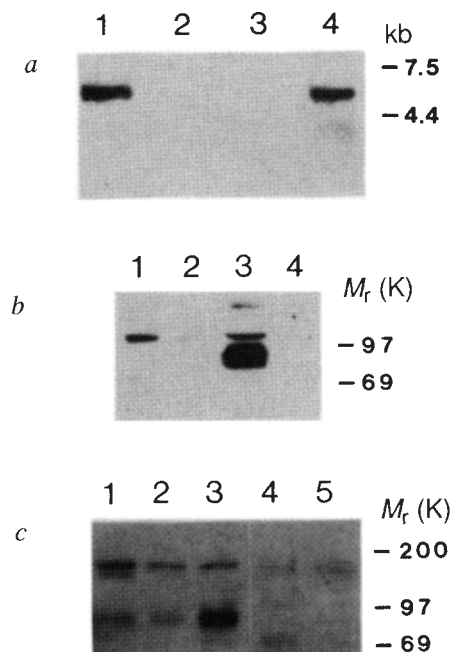


FIG. 4 *MDM2* expression. *a*, Northern blot analysis. RNA was separated by electrophoresis in a MOPS-formaldehyde gel and electrophoretically transferred to nylon filters. Transfer and hybridization were as described³⁰. RNA was hybridized to the *MDM2* fragment described in Fig. 3 legend. Total RNA (10 μ g) was derived, respectively, from two sarcoma cell lines (OsA-CL, lane 1 and RC13, lane 2) and the colorectal cancer cell line (CaCo-2) used to make the cDNA library (lane 3). Lane 4 contains 10 μ g polyadenylated CaCo-2 RNA. RNA sizes are shown on the right in kb. *b*, Western blot analysis of the sarcoma cell lines RC13 (lane 1), OsA-CL (lane 3), HOS (lane 4), and the carcinoma cell line CaCo-2 (lane 2). *c*, Western blot analysis of primary sarcomas. Lanes 1 to 3 contain protein from sarcomas with *MDM2* amplifications, and lanes 4 and 5 contain protein from sarcomas without *MDM2* amplification. Western blots using affinity-purified MDM2 antibody were performed with 50 μ g protein per lane as described³¹, except that the membranes were blocked in 10% non-fat dried milk and 10% goat serum, and secondary antibodies were coupled to horseradish peroxidase to allow chemiluminescent detection (Amersham ECL). MDM2 antibody was affinity-purified with a pATH-MDM2 fusion protein using methods described in ref. 31. Nonspecific reactive proteins of 75, 105 and 170K are seen in all lanes, irrespective of *MDM2* amplification. MDM2 proteins, of M_r 90K, were observed only in the MDM2-amplified tumours. Protein marker sizes are shown on the right.



(unpublished results with T. Tokino and D. Sidransky). The amplification of *MDM2* provides another provocative parallel between viral carcinogenesis and the naturally occurring genetic alterations underlying sporadic human cancer. □

Received 26 March; accepted 5 May 1992.

- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *Science* **253**, 49–53 (1991).
- Fakhrazadeh, S., Trusko, R. S. & George, D. *EMBOJ* **10**, 1565–1569 (1991).
- Hinds, P. W. *et al. Cell Growth Differ.* **1**, 571–580 (1990).
- Momand, J., Zambetti, G. P., Olson, D. C., George, D. L. & Levine, A. J. *Cell* (in the press).
- Mandahl, N. *et al. Genes Chrom. Cancer* **1**, 9–14 (1989).
- Turc-Carel, C. *et al. Cancer genet. Cytogenet.* **23**, 291–299 (1986).
- Meltzer, P. S. *et al. Cell Growth Differ.* **2**, 495–501 (1991).
- Lane, D. P. & Benichou, S. *Genes Dev.* **4**, 1–8 (1990).
- Werness, B. A., Levine, A. J. & Howley, P. M. *Science* **248**, 76–79 (1990).
- Ochman, H., Ajioka, J., Garza, D. & Hartl, D. in *PCR Technology: Principles and Applications for DNA Amplification* (ed. Erlich, H. A.) 105–111 (Stockton, New York, 1985).
- Tanaka, K. *et al. FEBS Lett.* **271**, 41–46 (1990).
- Pinna, L. A. *Biochim. biophys. Acta* **1054**, 267–284 (1990).
- Plashine, M. *Nature* **335**, 683–689 (1988).
- Harrison, S. C. *Nature* **353**, 715–719 (1991).
- Kinzler, K. W. *et al. Science* **251**, 1366–1370 (1991).
- Law, M. L. *et al. Ann. hum. Genet.* **50**, 131–137 (1986).
- Weiss, S. W. & Enzinger, F. M. *Cancer* **41**, 2250–2266 (1978).
- Malawar, M.M., Abelson, H. T. & Suit, H. D. in *Cancer: Principles and Practice of Oncology* (ed. DeVita, V. T., Hellman, S. & Rosenberg, S. A.) 1293–1342. (Lippincott, Philadelphia, 1985).
- Kinzler, K. W. *et al. Science* **236**, 70–73 (1987).
- Kinzler, K. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 1031–1035 (1986).
- Brodeur, G.M. & Seeger, R.C. *Cancer genet. Cytogenet.* **19**, 101–111 (1986).
- Gubler, U. & Hoffmann, B. J. *Gene* **25**, 263–268 (1983).
- Elledge, S. J., Mulligan, J. T., Pamer, S. W., Spottswood, M. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1731–1735 (1991).
- Feinberg, A. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Dasmahaptra, B., Rozhon, E. J. & Schwartz, J. *Nucleic Acids Res.* **15**, 3933 (1987).
- El-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. *Nature Genet.* **1**, 45–49 (1992).
- Reed, K. C. & Mann, D. A. *Nucleic Acids Res.* **13**, 7207–7215 (1985).
- Fearon, E. R. *et al. Science* **247**, 49–56 (1989).
- Vogelstein, B. *et al. Cancer Res.* **47**, 4806–4813 (1987).
- Kinzler, K. W., Ruppert, J. M., Bigner, S. H. & Vogelstein, B. *Nature* **332**, 371–374 (1988).
- Kinzler, K. W. & Vogelstein, B. *Molec. cell. Biol.* **10**, 634–642 (1990).

ACKNOWLEDGEMENTS. We thank A. J. Levine for the information about MDM2–p53 interaction that stimulated this work, M. L. Law for somatic cell hybrids, S. Elledge for the lambda YES vector, and T. Gwiazda for preparation of the manuscript. This work was supported by the Preuss Foundation, the Clayton Fund, and grants from the National Institutes of Health.

Wild-type p53 activates transcription *in vitro*

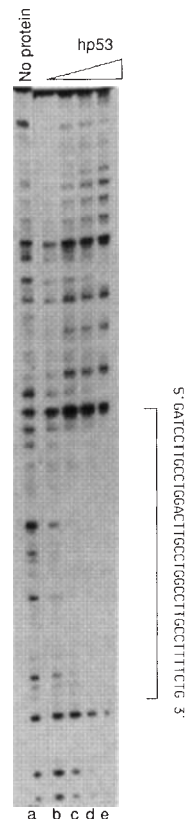
George Farmer, Jill Bargonetti, Hua Zhu, Paula Friedman, Ron Prywes & Carol Prives

Department of Biological Sciences, Columbia University, New York 10027, USA

THE p53 protein is an important determinant in human cancer and regulates the growth of cells in culture^{1–3}. It is known to be a sequence-specific DNA-binding protein^{4,5} with a powerful activation domain^{6–8}, but it has not been established whether it regulates transcription directly. Here we show that intact purified wild-type human and murine p53 proteins strongly activate transcription *in vitro*. This activation depends on the ability of p53 to bind to a template bearing a p53-binding sequence. By contrast, tumour-derived mutant p53 proteins cannot activate transcription from the template at all, and when complexed to wild-type p53, these mutants block transcriptional activation by the wild-type protein. Moreover, the simian virus 40 large T antigen inhibits wild-type p53 from activating transcription. Our results support a model in which p53 directly activates transcription but this activity can be inhibited by mutant p53 and SV40 large T antigen through interaction with wild-type p53.

A DNA-binding immunoassay has been used to screen human genomic clones and show that p53 binds specifically to a region upstream of the transcription start site for the human ribosomal gene cluster (RGC)⁴. We have confirmed and extended this observation by DNase I footprinting and shown that addition of immunopurified p53 to a DNA fragment containing the RGC

FIG. 1 The p53 protein binds specifically to a site in the human ribosomal gene cluster. DNA binding was assayed in 50- μ l volumes containing 40 mM creatine phosphate, pH 7.7, 4 mM ATP, 7 mM MgCl₂, bovine serum albumin (0.2 mg ml⁻¹), 0.5 mM dithiothreitol, 10 ng carrier plasmid (pAT153) and 10 fmol of 5' ³²P-labelled DNA fragment containing the ribosomal gene cluster (RGC) p53-binding site⁴ and either no protein (lane a) or increasing amounts of wild-type human p53 in increments of 15 ng up to 60 ng (lanes b–e). DNase I treatment of mixtures and processing of samples for electrophoresis on 8% polyacrylamide urea gels has been described⁵. Wild-type p53 was immunopurified from Sf27 cells expressing a recombinant baculovirus, pEV55hw, using the monoclonal antibody Pab421 crosslinked to Sepharose A¹⁴.



site leads to strong and specific protection of only the RGC region (Fig. 1). All tumour-derived mutant p53 proteins tested failed to protect this sequence (J.B. *et al.*, manuscript in preparation).

To determine whether p53 can activate transcription *in vitro*, we used as templates the plasmids fos1wt and fos1mt, which contain the human RGC p53 DNA-binding fragment or a mutated RGC fragment respectively (Fig. 2a). Three partially purified fractions from HeLa cell nuclear extracts were used as a source of transcription factors⁹. RNA products were analysed by S1 nuclease digestion using specific probes for each construct. Increasing amounts of p53 stimulated transcription from fos1wt (compare lanes 1–4 with lanes 5–8). These reaction mixtures also included a construct containing an abridged adenovirus major late promoter (pMLS; ref. 10) whose transcription was not significantly affected by p53.

The p53 protein activated transcription from another promoter as well (Fig. 2b). Plasmids containing either one or sixteen copies of the RGC site, or one mutant RGC site, inserted adjacent to the polyoma virus early promoter to create Pyl1wt, Pyl16wt and Pyl1mt, respectively, were used as templates in transcription reactions. We found that p53 activated transcription of constructs containing the wild-type RGC (lanes 1–9) but not the mutant RGC (lanes 10–12). Diagrams of the templates and the test probe used in these experiments are shown in Fig. 2c with the expected S1 nuclease products.

The high incidence of p53 gene mutations in cancer patients suggests that alteration of the normal function of p53 is an important part of the oncogenic process. Therefore it was of interest to examine whether tumour-derived mutant p53 proteins activate transcription. The mutant p53 proteins we chose are defective in both nonspecific and specific DNA binding^{4,5,11}, so providing an opportunity to confirm that p53 must bind DNA to activate transcription. We compared the ability of wild-type and two tumour-derived mutant p53 proteins with mutations at either amino acid 175 (His 175) or at amino acid 273 (His 273)

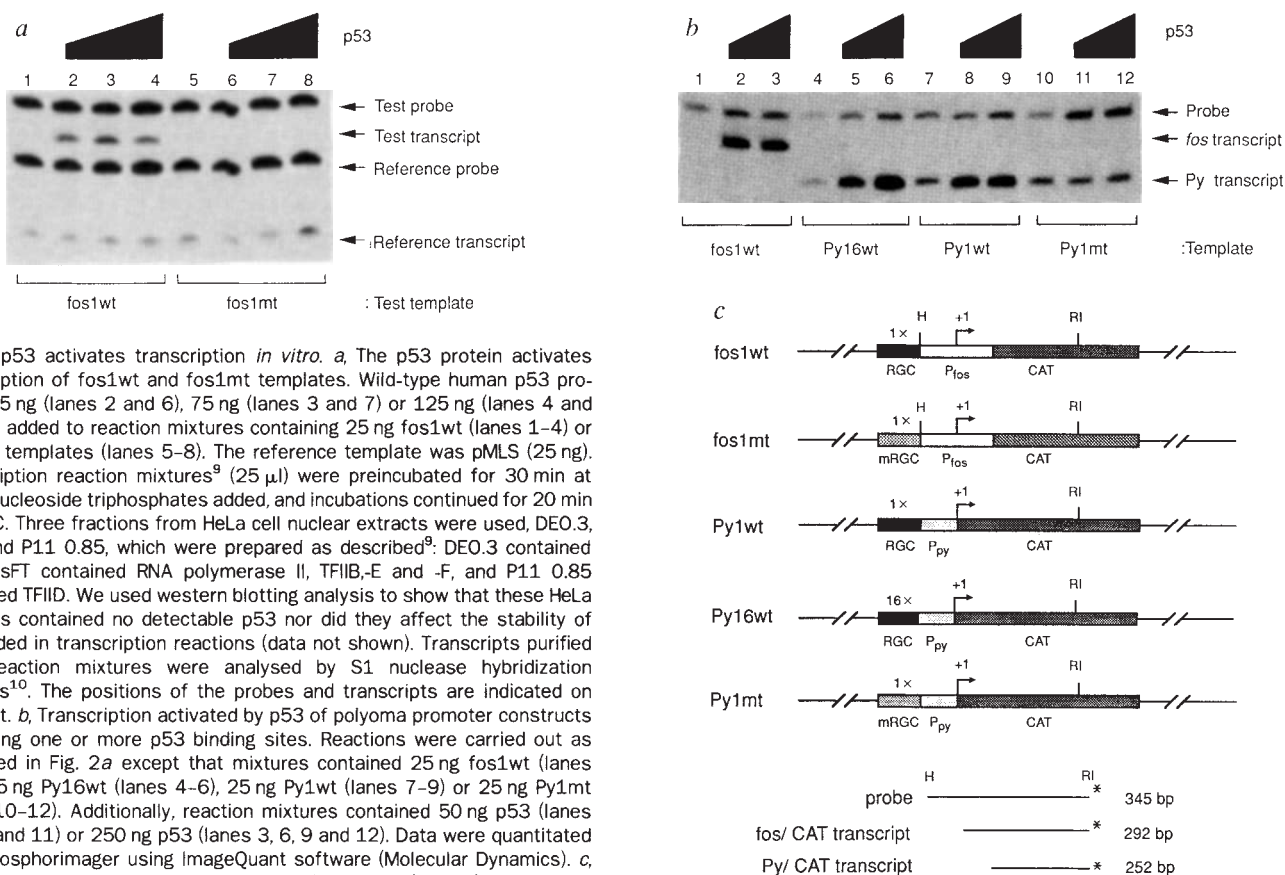


FIG. 2. p53 activates transcription *in vitro*. **a**, The p53 protein activates transcription of fos1wt and fos1mt templates. Wild-type human p53 proteins: 25 ng (lanes 2 and 6), 75 ng (lanes 3 and 7) or 125 ng (lanes 4 and 8) were added to reaction mixtures containing 25 ng fos1wt (lanes 1–4) or fos1mt templates (lanes 5–8). The reference template was pMLS (25 ng). Transcription reaction mixtures⁹ (25 μ l) were preincubated for 30 min at 30 °C, nucleoside triphosphates added, and incubations continued for 20 min at 30 °C. Three fractions from HeLa cell nuclear extracts were used, DEO.3, dsFT and P11 0.85, which were prepared as described⁹. DEO.3 contained TFIIA, dsFT contained RNA polymerase II, TFIIIB-E and -F, and P11 0.85 contained TFIID. We used western blotting analysis to show that these HeLa fractions contained no detectable p53 nor did they affect the stability of p53 added in transcription reactions (data not shown). Transcripts purified from reaction mixtures were analysed by S1 nuclease hybridization methods¹⁰. The positions of the probes and transcripts are indicated on the right. **b**, Transcription activated by p53 of polyoma promoter constructs containing one or more p53 binding sites. Reactions were carried out as described in Fig. 2a except that mixtures contained 25 ng fos1wt (lanes 1–3), 25 ng Py16wt (lanes 4–6), 25 ng Py1wt (lanes 7–9) or 25 ng Py1mt (lanes 10–12). Additionally, reaction mixtures contained 50 ng p53 (lanes 2, 5, 8 and 11) or 250 ng p53 (lanes 3, 6, 9 and 12). Data were quantitated by a Phosphorimager using ImageQuant software (Molecular Dynamics). **c**, Templates and test probe. To construct fos1wt, pFC53 (ref. 10) was cleaved with *Hind*III and *Kpn*I yielding a fragment containing *fos* sequences –53 to +42 fused to the bacterial chloramphenicol acetyl transferase (*CAT*) gene. This fragment was cloned downstream from the RGC sequences 5'-GTTGC-CTGGACTTGCCTGGCCTTTC-3' present in Bluescript SK P3P4 (ref. 4). fos1mt was constructed by ligating a double-stranded oligonucleotide containing a mutated RGC (mRGC) site: 5'-GTTTAAATGGACTTAAATGGCCTTAA-ATTTTC-3' into the *Xho*I site of pFC53. The mutated RGC site, whose sequence was determined by the results of methylation interference experiments, failed to bind specifically to p53 (ref. 4). Py1wt was constructed by inserting into Bluescript SK one RGC oligomer adjacent to polyoma sequences nucleotides 5,267 to 152) upstream of *CAT* from plasmid pPy01CAT

(ref. 29). Py16wt is the same as Py1wt but contains 16 concatemeric RGC sites. Py1mt contains 1 mutant RGC site as shown for fos1mt. Plasmids Py1wt, Py16wt and Py1mt were provided by S. Kern and B. Vogelstein. The test probe was a *Hind*III–*Eco*RI fragment of fos1wt end-labelled with ³²P at the *Eco*RI site. The higher levels of basal transcription of the polyoma constructs is probably due to nonspecific upstream initiation because this probe did not distinguish between nonspecific and specific initiation events from the polyoma promoters. Transcription initiation sites are indicated by arrows. P_{fos} and P_{py} refer to *fos* and polyoma promoters, H and RI to *Hind*III and *Eco*RI sites, respectively.

to activate transcription from fos1wt (Fig. 3). Each of these p53 mutations occurs frequently in human tumours². Unlike wild-type human p53, neither mutant protein is able to stimulate transcription.

One explanation for the oncogenic potential of mutant p53 proteins is that they can form hetero-oligomeric complexes with wild-type p53 and repress its normal function in growth regulation. Such wild-type/mutant p53 complexes have been identified^{12–14}. We therefore tested whether wild-type p53 could still activate transcription when complexed with either of our mutant p53 proteins. To generate these complexes, insect cells were co-infected with baculoviruses expressing wild-type murine p53 along with either wild-type human p53, His 175 or His 273 mutant human p53 proteins, respectively. When immunopurified p53 complexes from these cells were added to transcription reactions the results were striking: equivalent quantities of murine p53 either alone or in complex with wild-type human p53 activated transcription from fos1wt, but both murine mutant complexes were inactive at the highest concentration tested (Fig. 3b). Thus, not only are mutant p53 proteins themselves incapable of activating transcription, but they also prevent wild-type p53 from doing so.

In cells expressing the SV40 early region, a portion of the viral large T antigen is found in complex with p53 (ref. 15). As

T antigen is oncogenic^{15,16}, we investigated whether T antigen could affect the ability of p53 to activate transcription. We had previously shown that p53 and T antigen immunopurified from baculovirus insect cells form a complex *in vitro*¹⁷ and that T antigen blocks the ability of p53 to bind to sequences in the SV40 regulatory region⁵. The two proteins were preincubated either separately or together before addition to transcription reactions (Fig. 4). T antigen alone has no detectable effect on expression of fos1wt (lanes 7–9), and transcription of the pMLS reference plasmid is unaffected by T antigen (data not shown). By comparison, T antigen markedly inhibits the ability of p53 to stimulate expression from the template (Fig. 4a, compare lane 2 with lanes 4–6, and quantitation in Fig. 4b).

Our results demonstrate that p53 activates transcription from promoters bearing one or more p53-binding sites. Although the constructs used here are not physiological targets of p53, they do show that p53 can directly activate promoters. Our conclusions are supported by other experiments using the same polyoma promoter-containing constructs as we use here; these have shown that p53 activates expression from these constructs *in vivo*³⁰. Furthermore, the 5' flanking region of the murine muscle creatine kinase gene, reported to be stimulated by p53 in cell transfection experiments¹⁸, contains a strong binding site for wild-type p53 (ref. 31). Other genes whose activity is directly

activated by p53 should eventually be identified as a result of our observation.

The mutant p53 proteins tested here contain missense mutations that map to the central conserved region of the p53 protein¹⁹; their function may be to control the conformation of this region^{13,20}. We do not know whether this region also contains sequences that directly contact DNA. Our observation that complexes containing wild-type and mutant p53 proteins do not activate transcription is consistent with our results that such complexes do not bind the RGC sequence specifically (data not shown), and with a report that wild-type p53 assumes a mutant-like conformation when complexed with mutant p53 (ref. 21). Our wild-type murine/mutant human p53 complexes contain roughly equivalent concentrations of mutant human and wild-type proteins, which indicates that the active form of p53 could be a multimer and that even a few mutant p53 monomers are sufficient to inactivate the complex. Indeed, immunopurified p53 proteins can exist as tetramers and larger oligomeric structures (ref. 22 and P.F., J.B. and C.P., manuscript in preparation).

Can our results be reconciled with the rather broad negative

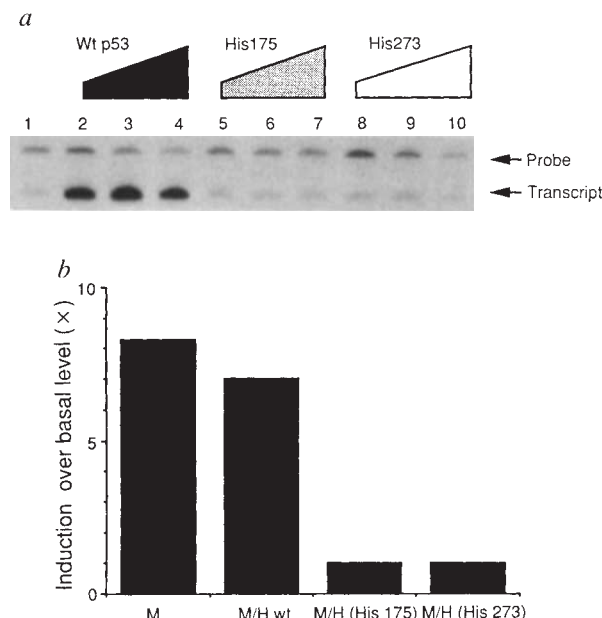


FIG. 3 Effect of mutant p53 proteins on transcription *in vitro*. *a*, Mutant p53 proteins do not activate transcription. Incubations were done as described in Fig. 2a, except that all reaction mixtures contained 25 ng fos1wt and 25, 75 or 125 ng of wild-type (wt; lanes 2, 3 and 4, respectively), 25, 75 or 125 ng of mutant (His 175) (lanes 5, 6 and 7, respectively), or 25, 75 or 125 ng mutant (His 273) (lanes 8, 9 and 10, respectively) p53 proteins. Mutants (His 175) and (His 273) were immunopurified from insect cells infected with recombinant baculoviruses⁵. *b*, Dominant negative effect of mutant p53 proteins. Reaction mixtures treated as for *a* with 25 ng fos1wt contained either no p53 or 150 ng wild-type murine p53 alone (M), or 250 ng murine p53 complexed with wild-type human (M/Hwt), mutant (His 175) (M/H His 175) or mutant (His 273) (M/H His 273)). RNA products were analysed by S1 nuclease digestion and quantitated using a Phosphorimager as described for *a*. Complexes of murine and human p53 proteins were obtained by co-infecting insect cells with recombinant baculoviruses expressing wild-type murine p53 and either wild-type human or mutant p53 baculoviruses. p53 proteins were immunopurified using p53-specific Pab421 protein A-Sepharose columns and were quantitated by colorimetric analysis and on silver-stained polyacrylamide gels. All complexes contained roughly the same amounts of murine p53 (data not shown) and in each murine-human p53 complex murine and human proteins were equally represented; these could be distinguished by their electrophoretic mobilities (data not shown). Each of the two p53 types must have been associated with the other because both murine and human p53 proteins in such complexes can be co-immunoprecipitated by either human-specific or mouse-specific anti-p53 monoclonal antibodies¹⁴.

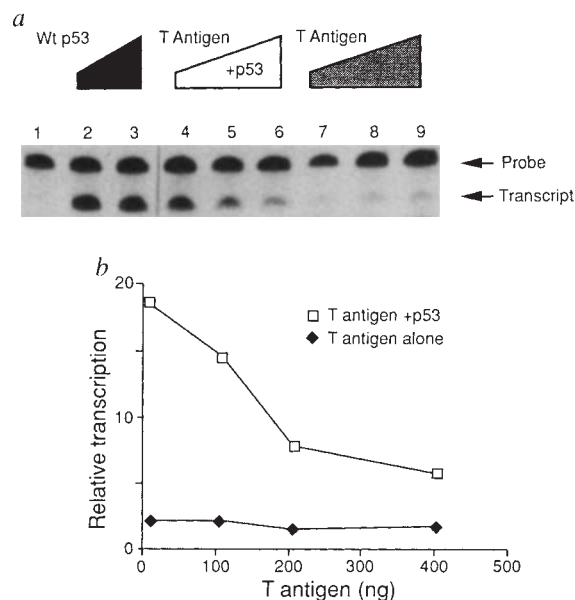


FIG. 4 SV40 large T antigen inhibits transcriptional activation by human p53 *in vitro*. *a*, Reaction mixtures were as described for Fig. 2a and contained 25 ng fos1wt. Before preincubation, 50 ng or 150 ng p53 were incubated for 10 min at 30 °C (lanes 2 and 3 respectively) while 50 ng p53 were incubated in the presence of 100, 200 or 400 ng T antigen (lanes 4, 5, and 6, respectively). T antigen (100, 200 or 400 ng) was incubated alone for the same time (lanes 7, 8 and 9, respectively). Proteins were then added to transcription reactions as described for Fig. 2a. SV40 T antigen was immunopurified from insect cells¹⁷. *b*, Graphical representation of results shown in *a* as quantitated by a Phosphorimager.

regulatory effect of wild-type p53 in repressing several promoters²³⁻²⁵? First, there could be one or more transcriptional repressors whose synthesis may be activated by p53. Second, the powerful activation domain of p53 when unattached to the template may prevent activation by other less effective factors, as in transcriptional 'squenching'²⁶; this would be more evident when p53 is overexpressed, a situation common to all reports describing its regulatory effects in cells. In our experiments, however, promoters lacking p53 binding sites were not repressed by addition of p53. Third, p53 may repress transcription either when its binding site is in the context of additional promoter elements, or when it binds to sequences other than those tested here.

We have shown that SV40 T antigen prevents p53 from activating transcription, which is consistent with a proposal that when transforming proteins encoded by different tumour viruses bind to p53, it is inactivated^{3,15}. For example, the E6 protein encoded by oncogenic strains of human papillomavirus targets p53 for rapid degradation^{27,28}. As T antigen prevents p53 from binding specifically to sites in both viral⁵ and cellular DNA (J.B. *et al.*, manuscript in preparation), this may be how p53 is prevented by T antigen from activating transcription. In view of the fact that neither tumour-derived mutant p53 proteins alone nor wild-type p53 when complexed with T antigen can bind DNA and activate transcription, it will be important to identify the genes that bind to and are stimulated by p53, as well as the factors with which p53 interacts to affect transcriptional activation. □

Received 14 February; accepted 14 May 1992.

1. Lane, D. P. & Benichou, S. *Genes Dev.* **4**, 1-8 (1990).
2. Hollstein, M., Sidransky, B., Vogelstein, B. & Harris, C. C. *Science* **253**, 49-53 (1991).
3. Levine, A. J., Momand, J. & Finlay, C. *Nature* **351**, 453-456 (1991).
4. Kern, S. E. *et al. Science* **252**, 1708-1711 (1991).
5. Bargonetti, J., Friedman, P. N., Kern, S. E., Vogelstein, B. & Prives, C. *Cell* **65**, 1083-1091 (1991).

6. Fields, S. & Jang, S. K. *Science* **249**, 1046–1049 (1990).
7. Raycroft, L., Wu, H. Y. & Lozano, G. *Science* **249**, 1049–1051 (1990).
8. O'Rourke, R. W. *et al.* *Oncogene* **5**, 1829–1832 (1990).
9. Zhu, H., Roy, A. L., Roeder, R. G. & Prywes, R. *New Biologist* **3**, 455–464 (1991).
10. Prywes, R., Fisch, T. M. & Roeder, R. G. *Cold Spring Harbor Symp. quant. Biol.* **53**, 739–748 (1988).
11. Kern, S. E. *et al.* *Oncogene* **6**, 131–136 (1991).
12. Martinez, J., Georgoff, I., Martinez, J. & Levine, A. J. *Genes Dev.* **5**, 151–159 (1991).
13. Milner, J., Medcalf, E. A. & Cook, A. C. *Molec. cell. Biol.* **11**, 12–19 (1991).
14. Friedman, P. N., Kern, S. E., Vogelstein, B. & Prives, C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9275–9279 (1990).
15. Levine, A. J. *Virology* **177**, 419–426 (1990).
16. Fried, M. & Prives, C. *Cancer Cells* **4**, 1–16 (1986).
17. Wang, E., Friedman, P. N. & Prives, C. *Cell* **57**, 379–392 (1989).
18. Weintraub, H., Hauschka, S. & Tapscott, S. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4750–4751 (1991).
19. Soussi, T., Caron de Fromental, C. & May, P. *Oncogene* **5**, 945–952 (1990).
20. Gannon, J. V., Greaves, R., Iggo, R. & Lane, D. P. *EMBO J.* **9**, 1595–1602 (1990).
21. Milner, J. & Medcalf, E. A. *Cell* **65**, 765–774 (1991).
22. Stenger, J. E. *et al.* *Molec. Carcinogenesis* **5**, 102–106 (1992).
23. Mercer, W. E., Shields, M. T., Lin, D., Appella, E. & Ullrich, S. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1958–1962 (1991).
24. Santhanum, U., Ray, A. & Sehgal, P. B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7605–7609 (1991).
25. Ginsberg, D., Mechta, F., Yaniv, M. & Oren, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9979–9983 (1991).
26. Plashne, M. *Nature* **335**, 683–689 (1988).
27. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. *Cell* **63**, 1129–1136 (1990).
28. Crook, T., Tidy, J. A. & Vousden, K. H. *Cell* **67**, 547–556 (1991).
29. Murakami, Y., Asano, M., Satake, M. & Ito, Y. *Oncogene* **5**, 5–14 (1990).
30. Kern, S. E. *et al.* *Science* **256**, 827–830 (1992).
31. Zambetti, G. P. *et al.* *Genes Dev.* (in the press).

ACKNOWLEDGEMENTS. We thank B. Vogelstein and S. Kern for sharing information and reagents, and E. Freulich for technical assistance. This work was supported by grants from the NIH (C.P.), March of Dimes Foundation and Searle Scholars Program (R.P.), and the Damon Runyon Fund (J.B.).

A new approach to protein fold recognition

D. T. Jones*†, W. R. Taylor† & J. M. Thornton*

* Biomolecular Structure and Modelling Unit,
Department of Biochemistry and Molecular Biology,
University College, Gower Street,
London WC1E 6BT, UK

† Laboratory of Mathematical Biology, National Institute for Medical Research,
The Ridgeway, Mill Hill, London, NW7 1AA, UK

THE prediction of protein tertiary structure from sequence using molecular energy calculations has not yet been successful; an alternative strategy of recognizing known motifs¹ or folds^{2–4} in sequences looks more promising. We present here a new approach to fold recognition, whereby sequences are fitted directly onto the backbone coordinates of known protein structures. Our method for protein fold recognition involves automatic modelling of protein structures using a given sequence, and is based on the frameworks of known protein folds. The plausibility of each model, and hence the degree of compatibility between the sequence and the proposed structure, is evaluated by means of a set of empirical potentials derived from proteins of known structure. The novel aspect of our approach is that the matching of sequences to backbone coordinates is performed in full three-dimensional space, incorporating specific pair interactions explicitly.

In outline our method is simple. A library of different protein folds is derived from the database of protein structures. In our case, the library contained all the unique, moderately well resolved chains (sequence identity < 30%, resolution ≤ 2.8 Å) in the July 1991 release of the Brookhaven database⁵, totalling 102 chains. Each fold is considered as a chain tracing through space; the original sequence being ignored completely. The test sequence is then optimally fitted to each library fold (allowing for relative insertions and deletions in loop regions), with the 'energy' of each possible fit (or threading) being calculated by summing the proposed pairwise interactions. The library of folds is then ranked in ascending order of total energy, with the lowest energy fold being taken as the most probable match.

In previous work, the difficult problem of optimizing the threading of the test sequence onto a structure with respect to the detailed pairwise interactions has been avoided by matching at the sequence level. At its most basic, this involves matching the sequence of the given fold with the test sequence, scoring each residue–residue match by means of a score matrix such as the Dayhoff matrix⁶. But there are now many examples of proteins exhibiting high structural similarity yet little or no similarity in their sequences (sequence identity < 15%). In view of this, several groups have attempted to match sequences to folds by describing the fold not in terms of its amino-acid sequence, but in terms of the environment of each residue location in the structure^{2–4}. The environment (for example, the local secondary structure and solvent accessibility) of a particular residue tends to be more highly conserved than the

identity of the residue itself, and so methods that match each residue in the test sequence to the environments of each residue in a protein fold are able to detect more distant sequence–structure relationships than purely sequence-based methods. Finkelstein and Reva attempted to take pairwise interactions into account by addressing the problem of fitting a sequence onto idealized lattice models of 8-stranded β -sandwich folds using an iterative procedure^{3,7}. We have used a dynamic programming-based algorithm^{8,9} capable of optimizing pairwise interactions, which was originally applied to the problem of structural comparison. This algorithm uses a standard sequence alignment method to optimize the threading of the sequence onto the structure around each residue in turn, finally computing the best threading through the whole structure by means of a shortest-path algorithm.

To evaluate the energy of a sequence in a particular conformation we need a set of potentials for residue interactions that do not require explicit modelling of all side-chain atoms. Previous work¹⁰ has shown that classical potentials (for example, CHARMM¹¹) cannot identify proteins that have been folded into non-native conformations. For these reasons we use a set of knowledge-based potentials and explicitly consider the degree of residue solvation, both of which do in fact identify such misfolded proteins^{12,13}. In particular, we use a set of pairwise potentials similar to those described by Sippl¹⁴ which are derived from a statistical analysis of known protein structures (see Fig. 1 legend for details). For a given pair of atoms, a given residue sequence separation and a given interaction distance, these potentials provide a measure of pseudo-energy, which relates to the probability of observing the proposed interaction in native protein structures. By dividing the empirical potentials into sequence separation ranges, specific structural significance may be tentatively conferred on each range. For instance, the short-range terms predominate in the matching of secondary structural elements. By threading a sequence segment onto the template of an α -helical conformation and evaluating the short-range potential terms, the probability of the sequence folding into an α -helix may be evaluated. In a similar way, medium-range terms mediate the matching of super-secondary structural motifs, and the long-range terms, the tertiary packing. Some sample potentials are shown in Fig. 1a–d.

Our medium and long-range pairwise potentials differ from those proposed in ref. 14 in that interactions beyond 10 Å are ignored. These interactions are not residue-specific and are determined simply by solvation effects. In place of these long-distance terms, we substitute a 'solvation potential'. This potential measures the propensity of each amino-acid type for a certain degree of solvation, approximated by the residue solvent-accessible surface area.

Many studies have shown that only the cores of distantly related structures are conserved, therefore in calculating the energy of a given threading we ignore all pairwise terms involving loop residues. Loop positions are evaluated by the solvation potential alone, which takes into account the tendency for loop regions to be exposed to solvent.

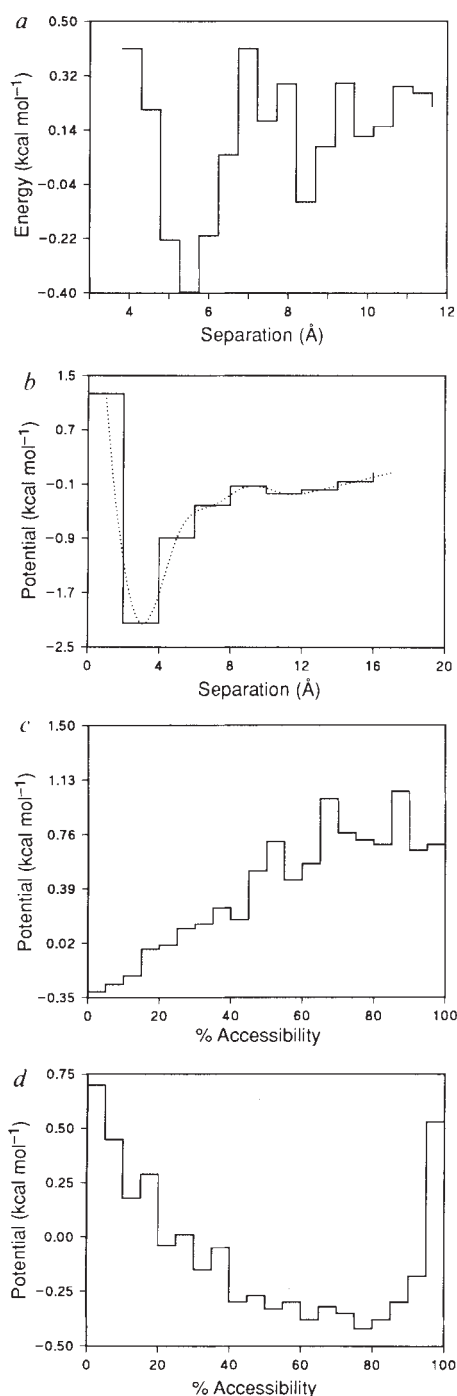


FIG. 1 Samples of the statistically derived potentials are shown. *a*, Short-range ($k=3$) Ala-Ala $C\beta \rightarrow C\beta$ interaction. Low-energy states are observed for distances around 6 Å, corresponding mainly to α -structure, and 9 Å, corresponding mainly to β -structure. *b*, Long-range ($k > 30$) Cys-Cys $C\beta \rightarrow C\beta$ interaction. The most significant energy minimum around 4 Å corresponds to disulphide bridge formation. *c*, Solvation potential for leucine, and *d*, solvation potential for glutamic acid. *e*, Threading histogram for the C-terminal ribosomal protein fragment, 1CTF. All possible threadings of the CTF sequence on the CTF structure were computed (secondary structure gaps disallowed) and the energies of each threading calculated. The native threading is indicated, and was found to be the lowest energy threading. METHODS. The calculation of pairwise pseudo-energy terms has been described¹⁴. For specified atoms ($C\beta \rightarrow C\beta$ for example) in a pair of residues *ab*, topological level (sequence separation) *k* and distance interval *s*, the potential is given by the following expression

$$\Delta E_k^{ab} = RT \ln [1 + m_{ab}\sigma] - RT \ln \left[1 + m_{ab}\sigma \frac{f_k^{ab}(s)}{f_k(s)} \right]$$

where m_{ab} is the number of pairs *ab* observed at topological level *k*, σ is the weight given to each observation, $f_k(s)$ is the frequency of occurrence of all residue pairs at topological level *k* and separation distance *s*, and $f_k^{ab}(s)$ is the equivalent frequency of occurrence of residue pair *ab*. *RT* is taken as 0.582 kcal mol⁻¹. Short- (sequence separation, $k \leq 10$), medium- ($11 \leq k \leq 30$) and long- ($k > 30$) range potentials have been calculated between the following atom pairs: $C\beta \rightarrow C\beta$, $C\beta \rightarrow N$, $C\beta \rightarrow O$, $N \rightarrow C\beta$, $N \rightarrow O$, $O \rightarrow C\beta$ and $O \rightarrow N$. Similarly, the solvation potential for an amino-acid residue *a* is defined as

$$\Delta E_{solv}^a(r) = -RT \ln \left[\frac{f^a(r)}{f(r)} \right]$$

where *r* is the % residue accessibility (relative to residue accessibility in GGXGG fully extended pentapeptide), $f^a(r)$ is the frequency of occurrence of residue *a* with accessibility *r*, and $f(r)$ is the frequency of occurrence of all residues with accessibility *r*. Residue accessibilities were calculated using the program DSSP¹⁹ applied to Brookhaven coordinate files. For multimeric proteins, only the chains explicitly included in the coordinate files were taken into account.

An obvious first test of these potentials was to attempt to thread a sequence onto its own native structure. Taking a number of small structures, for which it was practical to evaluate every possible threading (disallowing gaps in regions of regular secondary structure), we have found that the native threading of a sequence onto its own structure is usually found to be the lowest energy threading. As an example, the native threading histogram for the C-terminal ribosomal protein fragment (CTF) is shown in Fig. 1*e*. One small protein for which the native threading does not have the lowest evaluated energy is crambin, but this is attributable to the fact that this protein is not soluble in water, and consequently the solvation effects are not correctly modelled by our solvation potential.

To demonstrate the capability of our method for recognizing protein folds and generating an accurate sequence-structure

alignment, we consider here the example of C-phycoerythrin. The striking feature of the chain fold of C-phycoerythrin is that the globular portion (helices A-H) closely resembles the globin fold¹⁵. Despite the similarity in fold, the sequence homology between the globins and C-phycoerythrin is very low, with only 14 identities between the 174 β -chain residues of C-phycoerythrin and sperm whale myoglobin. So far, sequence analysis methods have proved unable to detect the globin fold in C-phycoerythrin. For example, despite success in constructing templates to select almost every available globin sequence, it has not been possible to match these templates against the phycoerythrins¹⁶. Using the optimal threading algorithm, the C-phycoerythrin sequence was threaded on each member of the library of protein folds to find its most compatible fold. The two lowest-energy folds were found to be sea hare myoglobin (-451 kcal mol⁻¹) and midge

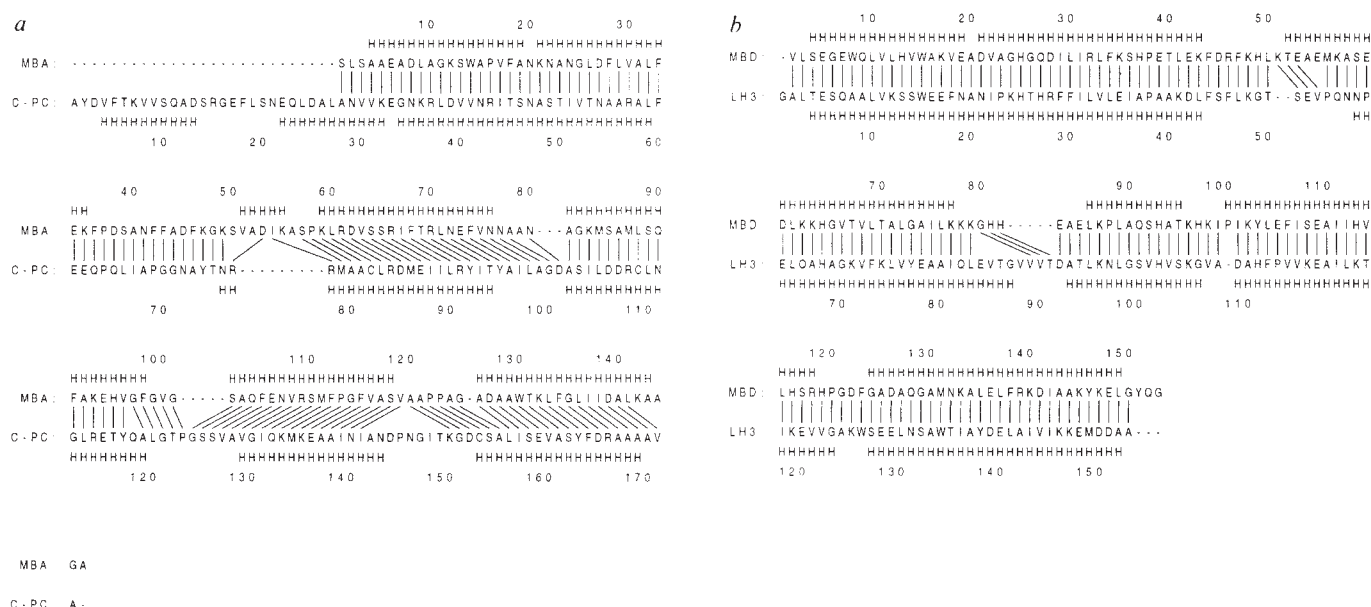


FIG. 2 *a*, Alignment of sea hare myoglobin (1MBA) with C-phycocyanin (C-PC) β -chain from *Mastigocladus laminosus* (SWISSPROT code PHCB\$MASLA), found by optimal threading. Author-assigned secondary-structure codes are shown. The alignment is compared to the structurally determined alignment by Pastore and Lesk, where lines are drawn between structurally equivalent

residue pairs as determined in the reference alignment¹⁷. *b*, Optimal threading of yellow lupin leghaemoglobin (LH3) on the structure of sperm whale myoglobin (Brookhaven code 1MBD). Alignment of protein structures is compared with the structural alignment obtained using the program SSAP⁹.

erythrocyruorin ($-356 \text{ kcal mol}^{-1}$) followed by several other all- α protein folds. Figure 2*a* shows the alignment corresponding to the optimal threading of the C-phycocyanin β -chain sequence onto the best matching fold (sea hare myoglobin). For comparison, the optimal threading alignment of myoglobin and leghaemoglobin is shown in Fig. 2*b*. In terms of sequence, myoglobin and leghaemoglobin are only distantly related (17%

sequence identity), but their structural similarity is much higher than in the case of phycocyanin and myoglobin, leading to a relatively unambiguous alignment. It should be borne in mind that even the structural alignment of phycocyanin and myoglobin is uncertain¹⁷. The fact that the optimal threading algorithm finds sea hare myoglobin to be the best model for C-phycocyanin is in accordance with a report¹⁷ in which the

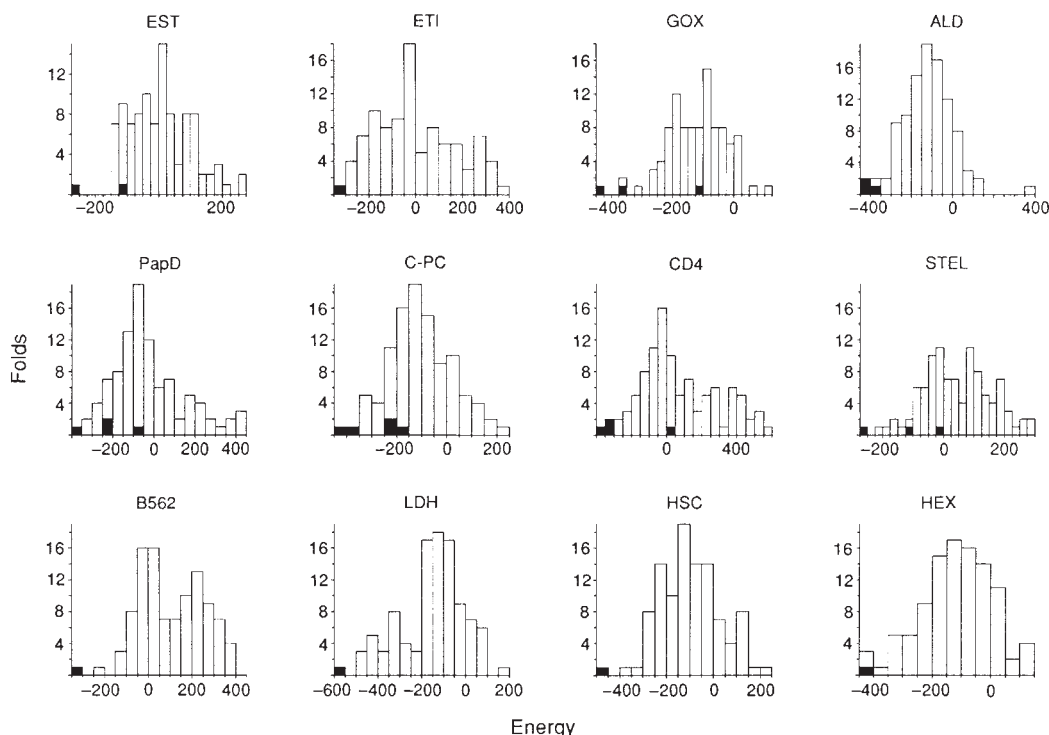
TABLE 1 Summary of trial fold-recognition searches

Test protein	Source	Fold	Best match	ΔE	% Sequence identity	Matches
C-phycocyanin β (C-PC)	Red algae	Globin	1MBA	101	7	1, 2, 9, 18, 25
Glycolate oxidase (GOX)	Spinach	TIM barrel	1WSY(A)	52	10	1, 3, 49
Muscle aldolase (ALD)	Human	TIM barrel	4XIA(A)	80	6	1, 2, 3
Lactate dehydrogenase (LDH)	Dogfish	Rossmann	4MDH(A)	87	15	1*
Elastase (EST)	Pig	Trypsin	4PTP	110	35	1, 14
CD4	Human	Ig	2FB4(H)	87	10	1, 2, 31
Stellacyanin (STEL)	Varnish tree	Cu binding	2AZA(A)	18	14	1, 6, 20
Cytochrome B562 (B562)	<i>E. coli</i>	4-helix bundle	2MHR	78	6	1
Trypsin inhibitor DE-3 (ETI)	Kaffir tree	Interleukin 1 β	111B	14	5	1
PapD-chaperonin	<i>E. coli</i>	Ig	2FB4(L)	64	15	1, 5, 9, 35
70K, Heat-shock cognate (HSC)	Cow	Actin	1ATN(A)	94	9	1
Hexokinase B (HEX)	Yeast	Actin	1ATN(A)	0	12	1

In each case the database included 102 protein chains, except where the test protein was itself in the database, in which case it was excluded. Templates for each chain were constructed as described in the text, with residues not in helices or strands (as calculated by DSSP¹⁹) assigned as loop residues. For the 70K heat-shock cognate protein and hexokinase searches, the coordinates for actin were also included (coordinates deposited under the code 1ATN, but not yet released). Proteins with >25% sequence identity to the test protein were also excluded from the calculation of potentials. The pairwise and solvation terms were summed and stored separately, and standard deviations ($s.d._{pair}$ and $s.d._{solv}$) for the two contributing factors calculated over the set of 102 folds. To balance the contributions of the pairwise and solvation terms, the final energy was taken as $E = E_{pair} + WE_{solv}$, where $W = (s.d._{pair}/s.d._{solv})$. The 'confidence' of the match (ΔE) is given in terms of the absolute energy difference between the top scoring fold and the next highest scoring, different, fold. The 'best match' column gives the Brookhaven ID of the best matching chain fold (including chain identity where appropriate), along with the sequence identity between the best matching chain and the test protein. Positions in the sorted list of threading energies of similar folds are also shown. A constant set of alignment parameters (gap penalty for example) was used for all databank searches shown. Typical execution times for a single search of 102 chains are around 100 minutes on a Unix workstation (Solbourne 5/602). The 102 chains used were as follows: 351C, A256B, 2AAT, 1ABP, A5ACN, 8ADH, 3ADK, A8ATC, A2AZA, 3BLM, 1BP2, 2CA2, A7CAT, 1CC5, 1CCR, A2CCY, 1CD4, 2CDV, 3CLA, 2CNA, 14CPA, 5CPA, 2CPP, 4CPV, 1CRN, 2CRO, E1CSE, 11CSE, 1CTF, 1CY3, 2CYP, 3DFR, A4DFR, A1DHF, 1ECA, E2ER7, H2FB4, L2FB4, 1FD2, 1FX1, 3FXC, 4FXN, A3GAP, 2GBP, 1GCR, 01GD1, 3GRS, A3HHB, 1HP, A2HLA, B2HLA, 1HOE, 111B, 3ICB, 3ICD, 1L01, 2LBP, 6LDH, 1LH1, 31LRD, A2LTN, 1LZ1, 1MBA, 1MBD, A4MDH, 2MHR, 2OVO, A2PAB, 9PAP, 1PAZ, 1PCY, A1PFK, 3PGK, 3PGM, 1PHH, 5PTI, 4PTP, 1RHD, 2RHE, 2RNT, 7RSA, 4RXN, 2SGA, 14SGB, 1SN3, 2SNS, O2SOD, 2SSI, 2STV, 1TGS, E2TMN, 4TNC, A1TNF, 1UBQ, 1UTG, A9WGA, R2WRP, A1WSY, B1WSY, A4XIA, A1YPI.

* Other topologically similar (yet structurally different) parallel $\alpha\beta$ folds were positioned at 3, 7, 11, 12, 13, 17, 19, 31, 34, 82.

FIG. 3 For a number of test cases (see Table 1) the histogram of energies for optimally threading onto each of the 102 folds is given. In each histogram, the positions of folds expected to match the given sequence (that is, those folds similar to the known fold of the test sequence) are shown as filled bars. For example, in the case of LDH (lactate dehydrogenase), the expected match in the database of folds is MDH (malate dehydrogenase). This match is shown as a single filled bar representing an energy of $-577 \text{ kcal mol}^{-1}$, an energy which is lower than that achieved by any other fold. As noted in the text, in some cases expected folds are apparently not detected. This occurs for two reasons: either the expected structures are not sufficiently similar to the native fold, or the optimization method does not succeed in producing a satisfactory alignment. The C-PC results demonstrate the former case. A number of unrelated highly helical proteins (carp parvalbumin and T4 lysozyme, for example) score better than the low-scoring globins. The worst-scoring globin is in fact human haemoglobin, in which case the poor score is due not only to the substantial secondary structural shifts relative to C-PC, but also to the fact that the calculated accessibilities are for the complete tetramer. The second situation arises in the results for GOX, where the algorithm fails to find an optimal threading of GOX onto XIA (xylose isomerase), resulting in an unexpectedly poor score. Of particular note in the results shown are the $(\alpha\beta)_8$ (TIM) barrel and trypsin inhibitor DE-3 examples. The degree of sequence homology between different $(\alpha\beta)_8$ barrel enzyme families and between trypsin inhibitor DE-3 and interleukin-1 β is extremely low (5–10%). As a consequence of this,



helix geometry of this globin was found to be closest to that of C-phycoerythrin. Not only has the method correctly identified its globin fold, but has accurately located it in the C-phycoerythrin sequence and has generated an alignment close to that obtained by careful structural alignment. It is clear that the method has identified the related folds in the database. It should be emphasized that no specific sequence information was used in the threading process: the structure was considered only as a chain of anonymous placeholders onto which the given sequence is threaded.

The results of other trial searches using the method of optimal sequence threading are shown in Table 1 and Fig. 3. From Fig. 3 it is apparent that in some cases expected matches are far

again, sequence template methods have been unable to detect these folds. Also of note are the results for the 70K heat-shock cognate protein (HSC70), and yeast hexokinase B. The N-terminal ATPase fragment of the heat-shock cognate protein has an almost identical structure to that of actin, but the similarity between hexokinase and actin is more topological than at the level of specific structural interactions. The two degrees of similarity are borne out by the threading results for these proteins, in that although actin is the lowest energy fold for hexokinase, the separation between the actin fold and the next-best-matching fold (aspartate transcarbamylase, ATC) is almost zero ($0.1 \text{ kcal mol}^{-1}$); the rather weak structural similarity between hexokinase and actin would therefore appear to be just at the limits of our method. In contrast, the match between HSC70 and actin is clearly significant.

from the top of the list. On inspection it was found that in these cases the threading algorithm had clearly misaligned the proteins, and had failed to find a reasonable optimum of the objective function, although this could generally be corrected by adjusting the alignment gap penalty.

The method described here shows promise as a new means for sensitively recognizing protein folds, and it is evident from the results that new information beyond sequence similarity is being exploited here. We are now exploring the generation of model folds^{3,18}, to escape from the limitation of only being able to predict previously observed folds, and the incorporation of multiple sequence data (from aligned sequence families) in the recognition process. □

Received 5 February; accepted 21 May 1992.

1. Taylor, W. R. & Thornton, J. M. *J. molec. Biol.* **173**, 487–514 (1984).
2. Overington, J., Johnson, M. S., Sali, A. & Blundell, T. L. *Proc. R. Soc. B* **241**, 132–145 (1990).
3. Finkelstein, A. V. & Reva, B. A. *Nature* **351**, 497–499 (1991).
4. Bowie, J. U., Lüthy, R. & Eisenberg, D. *Science* **253**, 164–170 (1991).
5. Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535–542 (1977).
6. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. in *Atlas of Protein Sequence and Structure* Vol. 5 suppl. 3 345–352 (Natl. Biomed. Res. Fnd, Washington DC, 1978).
7. Thornton, J. M., Flores, T. P., Jones, D. T. & Swindells, M. B. *Nature* **354**, 105–106 (1991).
8. Taylor, W. R. & Orengo, C. A. *J. molec. Biol.* **208**, 1–22 (1989).
9. Orengo, C. A. & Taylor, W. R. *J. theor. Biol.* **147**, 517–551 (1990).
10. Novotny, J., Brucoleri, R. E. & Karplus, M. *J. molec. Biol.* **177**, 787–818 (1984).

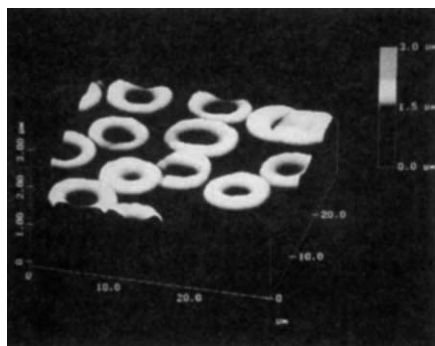
11. Brooks, B. *et al. J. comp. Chem.* **4**, 187–217 (1983).
12. Hendlich, M. *et al. J. molec. Biol.* **216**, 167–180 (1990).
13. Eisenberg, D. & McLachlan, A. D. *Nature* **319**, 199–203 (1986).
14. Sippl, M. J. *J. molec. Biol.* **213**, 859–883 (1990).
15. Schirmer, T., Bode, W., Huber, R., Sidler, W. & Zuber, H. *J. molec. Biol.* **184**, 257–277 (1985).
16. Bashford, D., Chothia, C. & Lesk, A. M. *J. molec. Biol.* **196**, 199–216 (1987).
17. Pastore, A. & Lesk, A. M. *Proteins* **8**, 133–155 (1990).
18. Taylor, W. R. *Prot. Engng* **4**, 853–870 (1991).
19. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).

ACKNOWLEDGEMENTS. We thank T. P. Flores, S. J. Hubbard, C. A. Orengo and M. B. Swindells for discussion, and K. C. Holmes for permission to use the coordinates for actin. D.T.J. acknowledges receipt of an SERC CASE studentship with the MRC.

Summertime specials

Featured this week — a hand-held laboratory computer that makes light work of calculating solutions, an extracellular matrix preparation for anchorage-dependent cells and software that simulates objects in real motion.

IMAGING Services is offering an **atomic force microscopy and scanning tunnelling microscopy service**, providing researchers with access to AFM/STM scans without having to purchase expensive equipment (*Reader Service No. 100*). Images are ob-



Atomic force microscopy image of red blood cells in air on glass.

tained (with the Nanoscope instrument from Digital Instruments) with three-dimensional resolution at nanometer scale. Scan sizes range from the atomic scale to $125 \times 125 \mu\text{m}$. Samples can be imaged in air or a variety of fluids. Biological applications of the AFM/STM include imaging of proteins, peptides or real-time biological processes.

Take the drudgery and potential for error out of making difficult solution calculations with the new Lab Partner Plus **hand-held laboratory computer** from Calculated Solutions (*Reader Service No. 101*). Routine calculations, such as the amount of solute or solvent needed for solutions or dilutions, and computations or conversion of any weight, volume, concentration, or activity to any other, can be made quickly. Even difficult calculations, such as mixtures involving radioactive materials, can be calculated at the touch of a button, the



Make no mistake with the Lab Partner Plus.

company says. An optional compact printer, which operates in tandem with Lab Partner Plus, provides a complete audit trail of all calculations. The Lab Printer automatically prints out the day, time and name of the program being used, the questions prompted by the computer, the user's answers to the prompts, and the answer(s) to the problems being solved.

A generic VAX/VMS version of Sweep, **virus-scanning software** from Sophos, is now available for Pathworks, Digital Equipment's VAX-based PC networking system (*Reader Service No. 102*). Sweep for Pathworks runs as a VAX/VMS process, checking DOS files held on the VAX server. It can run in the background, constantly checking DOS executable files and sounding the alarm if a virus is found. Sweep for Pathworks is updated every month as new viruses appear, with updates sent out on TK50 tape cartridges. Licences for Sweep for Pathworks are available on a 'per user' basis or a 'per VUP' (VAX unit of performance) basis. Prices start at £29.50 (UK) per user per annum, or £995 (UK) per VUP per annum, and include the PC version of Sweep.

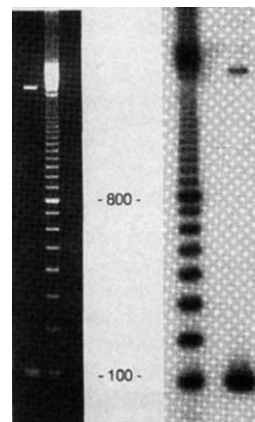
Molecular means

Novagen announces the introduction of the pCITE-2 series of vectors designed for **increased translational performance of RNA prepared from cloned DNA sequences** (*Reader Service No. 103*). The plasmids carry a segment of the encephalomyocarditis virus RNA 5' noncoding region, which functions as an internal entry point for initiation of translation by eukaryotic ribosomes. This cap-independent translation enhancer (CITE) sequence dramatically increases the *in vitro* translation efficiency of synthetic RNA by rabbit reticulocyte lysates and has a similar effect on transfected mammalian cells, Novagen says. The newly developed vectors now feature more cloning sites, an optional His-Tag leader for rapid affinity purification of translation products and an f1 origin of replication.

In addition to **custom DNA and RNA synthesis**, the Midland Certified Reagent Company offers a host of other services and products for the molecular biologist (*Reader Service No. 104*). Midland offers a choice of a standard custom DNA synthesis service, which includes deprotection and removal of ammonia and protecting groups by gel filtration, or a purified standard synthesis, which, in addition to the above, includes a

further purification step using anion-exchange HPLC, reversed-phase HPLC, polyacrylamide gel electrophoresis, or a combination of these methods. For researchers who need DNA in a hurry, Midland offers a 'next day' DNA delivery service. The company also offers polydeoxynucleotides prepared from high-purity deoxynucleoside triphosphates; polyribonucleotides built up from the monomeric ribonucleotides, adenylic, cytidylic, guanylic and/or uridylic acids; terminal deoxynucleotidyl transferase; DNA polymerase; reverse transcriptase/DNA polymerase substrates; oligodeoxythymidylates; oligodeoxyadenylates, oligodeoxycytidylates and oligodeoxyguanylates; linkers and adaptors. In addition, Midland has introduced a series of DNA probes for transcription factors.

The new **100-base-pair ladder** from Pharmacia P-L Biochemicals is designed for sizing polymerase chain reaction (PCR) products and other DNA fragments in the range of 100–2,000 base pairs (*Reader Service No. 105*). Each band of the ladder represents 100 bp. The ladder features a double-intensity band at 800 bp as a size reference for fast orientation.



On your marks.

Cambio is distributing the new Genomix **DNA extraction kit** from Talent, which allows researchers to extract DNA from human whole blood, white-cell nuclei, cultured cells or tissue samples using a method that provides large molecular weight DNA without shearing (*Reader Service No. 106*). The kit is designed to make DNA extraction more environmentally and user-friendly: no phenol is used and the researcher is protected from chloroform by a proteinaceous plug. In addition, any hazardous viral particles are inactivated during the first lysis step by a powerful denaturing cationic detergent. Using the kit, it is possible to extract about 30–40 micrograms of DNA per millilitre of blood (incubation with proteolytic enzymes is unnecessary). The

manual process has been automated by using Genomix with a new automatic DNA extraction machine called the Extragen 8C. According to Cambio, the instrument can perform eight sample extractions in 40 minutes and DNA is delivered at the end of the extraction process in solution.

Tools of the trade

A new line of products related to **eukaryotic topoisomerases** is now available from TopoGEN and should be of interest to researchers involved in anti-viral and anti-cancer drug design and discovery (*Reader*



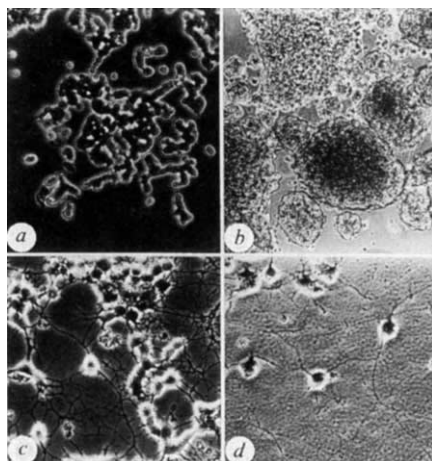
Topoisomerase-related products.

Service No. 107). The company offers purified human and calf type I or II topoisomerases, antibodies and DNA substrates for assaying topoisomerases. These products are available as single reagents or as components in assay kits and drug screening kits to facilitate detection of topoisomerases and discovery of novel inhibitors. Each kit includes buffers, control inhibitors, DNA substrates (kinetoplast or plasmid DNA) and protocols.

Becton Dickinson announces the availability of a new line of CAMFolio **monoclonal antibodies to human cell adhesion molecules** (*Reader Service No. 108*). The monoclonal antibodies, which are supplied as high-purity, ready-to-use formulations, are intended for use in applications such as adhesion blockade, functional studies, ELISAs and immunoprecipitation. The initial selection of 21 CAMFolio monoclonal antibodies includes portfolios of complementary molecules for research applications in cell biology, immunology, haematology and neurobiology. This portfolio approach is intended to facilitate experimental design in studies involving interactions of several cell adhesion molecules (for example, LFA-2/LFA-3 and ICAM-1/LFA-1 α in the migration of T lymphocytes across the blood-brain barrier: *Immunology Today* **12**, 277-282; 1991). The line-up includes anti-BB-1, anti-GMP-140, anti-gp150,95, anti-HCAM, anti-HNK-1, anti-HPCA-2, anti-ICAM-1, anti-integrin $\alpha 4$, anti-LECAM-1, anti-LFA-1 α , anti-LFA-1 β , anti-LFA-2, anti-LFA-3, anti-Mac-1, anti-N-CAM, anti-PECAM-1, anti-TCR α/β , anti-TCR γ/δ , CD5, CD22 and CD28. All are purified to 95 per cent purity as determined by SDS-polyacryl-

amide gel electrophoresis and are provided at a concentration of 1.0 mg ml⁻¹ with no added carrier proteins.

The new Biomatrix I **extracellular matrix preparation** from Biomedical Technologies is designed to promote the cellular attachment and differentiation of anchorage-dependent cell types (*Reader Service No. 109*). Biomatrix I is an extract from Engelbreth-Holm-Swarm tumours which contain all the components in basement membrane. These include laminin, type-IV collagen, entactin, heparin sulphate proteoglycan and others. BTI says that neurons, vascular endothelial cells, hair follicle cells and melanoma cells have been effectively grown on this preparation. Each lot of Biomatrix I is analysed for the attachment and differentiation of PC-12 cells in long-term culture. It is packaged in 1-ml (1 mg ml⁻¹) aliquots and costs \$100 (US) for 10 aliquots. The figure depicts PC-12 cells before (a) and 7 days after culture with



40 ng ml⁻¹ nerve growth factor (b,c,d). Plastic surfaces were either untreated (b) or treated (a,c,d) with Biomatrix I before cell plating. The Biomatrix I was either diluted and air dried (a,c) or undiluted and gelled (d).

All kitted out

Parathyroid hormone related peptide, a hormone that regulates calcium in human bones, may play an important role in the detection of cancer and treatment of osteoporosis. Its exact mechanism of action is not fully understood, but parathyroid hormone related peptide seems to activate osteoclasts, a type of cell that normally dissolves unwanted bone, transferring a large amount of calcium from bone to the bloodstream. It has been described in the literature that high concentrations of parathyroid hormone related peptide have been found in patients with tumours of the lung, breast, ovaries or bone marrow. The Nichols Institute has developed a two-stage **immunoradiometric assay for the measurement of parathyroid hormone related peptide**, which may be useful as a means of identifying early cancerous tumours (*Reader Service No. 110*).

The assay uses two affinity-purified polyclonal antibodies, each specific to well-defined regions on the parathyroid hormone related peptide molecule. One antibody, specific for the 60-72 amino acid region, is biotinylated, the other, directed towards the amino terminal (1-40 region), is iodinated for detection. Synthetic parathyroid hormone related peptide (1-86) is used for the standards. A sandwich complex forms in the presence of parathyroid hormone related peptide molecules that bridge these two antibodies. Avidin-coated beads are added to separate free from bound.

A large number integral membrane proteins are attached to the lipid bilayer of cell membranes through a covalent bond from the C-terminal amino acid of the polypeptide to the ethanolamine phosphate of a glycosylphosphatidylinositol (GPI) moiety. Such moieties, commonly referred to as GPI membrane anchors, are found on a variety of cell-surface glycoproteins, in many different cell types, and in species ranging from unicellular organisms to man. They have been implicated in transmembrane signalling and regulatory functions, and in a variety of disease states. Despite this ubiquity, the GPI membrane anchors have been characterized in only a few glycoproteins. Traditionally, most of the evidence implicating the presence of GPI anchors is their sensitivity to cleavage by GPI-specific phospholipase C. To provide even greater accuracy and confidence in analysis, GPI-phospholipase C sensitivity is only one of several criteria used in a new **glycobiology research kit** from Oxford GlycoSystems — the GPI anchor detection kit (*Reader Service No. 111*). The kit contains all the necessary reagents and materials to detect GPI membrane anchors in at least 5-10 samples. Protocols are included for samples such as cell lysates, pure glycoproteins and immunoprecipitates.

ADVERTISEMENT

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID

Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

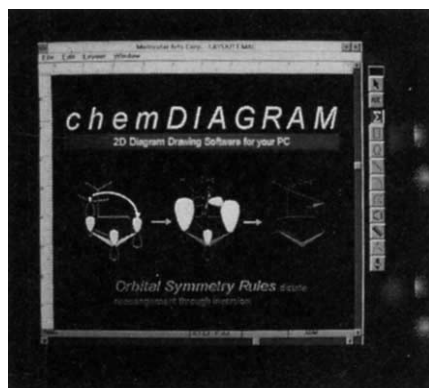
POB 126, D-8000 Munich 65, FRG

Reader Service No.21

Program profile

SigmaPlot 5.0 for DOS is the latest version of Jandel Scientific's graphing program, which can be used to create **three-dimensional graphs** (*Reader Service No. 112*). With version 5.0, users can create mesh and scatter plots in 3D with frame lines, drop lines, rotated labels and backplanes with colour and grids. The program now features hierarchical sub-menus and improved user prompts. In addition, creating a graph has been streamlined into a three-step procedure. A new feature of version 5.0 is the 'file merging' command, which takes a user-compiled list of SigmaPlot graphs and merges all the contents of each file onto the page, in a quick-and-easy step, Jandel says. SigmaPlot will also re-size and scale the merged graphs if this is requested by the user. Additional import options for Lotus, Symphony, Quattro and Microsoft Excel files have been added. SigmaPlot 5.0 for DOS costs DM 1,275 (Germany).

The latest software program to roll off the Molecular Arts production line is chemDIAGRAM, a PC-based software tool for drawing two-dimensional chemical diagrams (*Reader Service No. 113*). With chemDIAGRAM, users can create complex structure diagrams, move, scale, stack and rotate them, add labels and text, create illustrations to accompany research documentation and produce publication-quality output. The \$349 (US) program takes advantage of the Microsoft Windows graphical user interface and features a complete palette of tools for drawing diagrams, including 12 bond tools, 12 ring tools, 3 atom tools and 12 orbital tools. Other useful features include the 'snap to' drawing guides for atom positions, bond lengths, bond angles and labels. Users can also customize many of the diagram features, including bond widths, bond spacings, line weights, fonts, colours, bond stacking and many others. In addition to being a drawing program, there is a graphics drawing tool set, a text editor, special mathematic and symbol



2-D or not 2-D — see chemDIAGRAM.

libraries, and advanced page layout tools. Full support for cut-and-paste operations allows users to transfer illustrations into other Windows-compatible applications, such as word processors, spreadsheets and databases.

Interactive Physics II software from Knowledge Revolution is a complete **motion laboratory for the Macintosh** that simulates and measures objects in motion, driven by physical laws (*Reader Service No. 114*). With the program, users can draw and build any number of objects on a Macintosh screen, define motion parameters for each object (mass, elasticity, charge, velocity), set the environment (gravity, air resistance, electrostatics) of an experiment, and then immediately 'run' the experiment, simulating how the objects would interact in reality. The dynamic simulation engine mathematically creates smooth animation-like simulations, whose measurement data can be displayed simultaneously in graphical, meter or table format. The \$399 (US) Interactive Physics II program can be used to model simple events such as pendulums or objects in projectile motion; the same program can also create and run complex experiments such as solar system simulations and electrostatic/dynamic simulations. The software also supports Apple computer's System 7 and QuickTime.

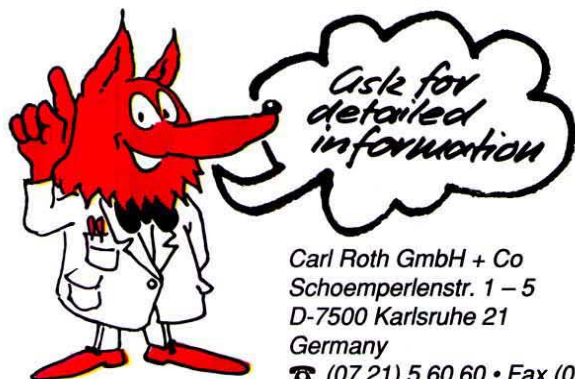
Handy hardware

The new **thin layer chromatography plate heater** from Camag, which holds TLC plates at a set temperature after derivatization reagent has been applied, is designed to maintain a uniform temperature over the entire surface of the heater (*Reader Service No. 115*). Camag says the ceramic hotplate is stable to all customary reagents and can be easily cleaned. Other features of the heater include a 200 × 200 mm heating area, digital display of set and actual temperatures, and a temperature range of 25–200 °C.

The German firm, H + P Labortechnik has introduced a line of **motorless magnetic stirrers** with 6, 15 or 60 stirring points (*Reader Service No. 116*). The version with 15 stirring points can be put to a wide variety of uses: it can take, for example, fifteen 250-ml beakers or also six 1000-ml Erlenmeyer flasks. The units are hermetically encapsulated in a stainless steel housing coated with a black synthetic material. This makes the units resistant to acids and alkalis and allows them to be immersed in water for cleaning. H + P Labortechnik says the units can even be used in incubators where there is a high level of atmospheric humidity.

Researchers needing to stir cultures or reagents under carefully controlled conditions may be interested in the new **SI 40 incubator with built-in magnetic stirrer** from Stuart Scientific (*Reader Service No. 117*). The 45-cm wide × 41.5-cm deep incubator houses a six-place magnetic stirrer that can accommodate vessels up to 280 mm in height (which includes culture flasks). Stirring speeds can be set to between 10 and 110 r.p.m. According to Stuart, electronic-feedback speed control ensures a gentle start to stirring and maintains the set speed even with changes in viscosity. The temperature can be varied from ambient + 5 °C to 65 °C. For added flexibility, the incubator can be controlled independently of the stirrer mechanism. □

ADVERTISEMENT



Carl Roth GmbH + Co
Schoemperlenstr. 1 – 5
D-7500 Karlsruhe 21
Germany
☎ (07 21) 5 60 60 • Fax (07 21) 56 06 49

Phenol ready-to-use solutions for DNA/RNA-separation

Roti®-phenol (equilibrated phenol)
Roti®-phenol / chloroform / isoamylalcohol

Chemicals + Labware
for Life Science

